Chemistry's understated majesty

The International Year of Chemistry is under way. Chemists should celebrate their discipline's past as the foundation of other fields, and face the future with increasing confidence.

"Chemistry can be a good and bad thing," said US comedian Adam Sandler. "Chemistry is good when you make love with it. Chemistry is bad when you make crack with it." The plentiful good points of chemistry are to be promoted to the public this year. Under the official United Nations banner of the International Year of Chemistry, chemists across the world are to join hands to celebrate their subject. Why 2011? Formally, it is to mark the 100th anniversary of Marie Curie receiving the Nobel Prize in Chemistry, for her discovery of radium and polonium. Informally, chemists are no doubt keen for their time in the sun. Last year was the Year of Biodiversity and 2009 was the Year of Astronomy. The International Polar Year stretched from 2007 into 2008, and 2005 was the Year of Physics. Chemistry's time has surely come — it is 350 years, after all, since Robert Boyle published *The Sceptical Chymist* and put the discipline firmly on the modern scientific map (see page 30).

It is regrettable for chemistry's champions that its iconic figures have to be shared with other disciplines. In the English-speaking world, Boyle is perhaps most commonly associated with his law relating the pressure and volume of a gas, and so with physics. And he is not the only shared standard bearer. Even the organizers of this year's chemistry extravaganza are trading on the reputation of someone whose fame relies heavily on her fatal fascination with radioactivity, for which Curie was awarded the 1903 Nobel Prize in Physics.

This situation is even more acute in modern research. Chemistry is often central, with principles and discoveries that enable work in other subjects. Its ability to react and rearrange matter for applications such as energy storage, new materials and more efficient industrial processes is vital for modern technology. Yet often, other disciplines such as materials science emerge as the public faces of such successes.

UNSUNG HEROES

Nature drew attention to this syndrome in 2001 (*Nature* **411,** 399; 2001). The "lack of an accurate and identifiable chemistry 'brand' means that the discipline is easily misunderstood", we wrote, "and those working in it are frequently under-appreciated". A decade on, this has become a popular cause for those who seek to promote the benefits of the chemical sciences. The organizers of this year's celebration, for example, want to "improve the understanding and appreciation of chemistry by the public" and "promote the role of chemistry in contributing to solutions to global challenges".

There is no question that chemistry's important advances often lack the widespread recognition they deserve. What can be done to change this? Although chemists are right to stress that their work underpins much of modern science, those foundations are rarely acclaimed. But perhaps chemistry has less a problem of public image than a lack of desire to assert itself. For beneath the utilitarian way that chemistry feeds into other disciplines, it has a rare and wonderful majesty on its own terms. Perhaps this is satisfaction enough for many. Witness this week's issue of *Nature*: alongside varied perspectives about chemistry and its future (see page 23), we publish three important papers across the field, in supramolecular chemistry, organic chemistry and biochemistry (see pages 72, 76 and 116). And as the year unfolds, we intend to publish more outstanding research from the molecular sciences, as well as a series of reviews on some of the most compelling topics in the field.

Nevertheless, as science funding becomes more competitive and is judged on visible results, the organizers of the chemistry year are right to seek credit where it is due. Perhaps the spotlight will shift towards chemists as boundaries between fields continue to blur. As biologists, for instance, zoom in on the action of molecules to probe physiological and cellular processes, including how cells communicate, they are effectively working as chemists — albeit chemists who work with oversized molecules.

> *"There is no shortage of problems to which chemists can contribute solutions."*

There are other ways to boost the profile of chemistry. In biology papers, and many synthetic-chemistry papers, key basic-chemistry references — descriptions of synthesis and characterization of individual compounds — are often relegated to supplementary information, where they can languish unnoticed and uncredited. Chemists often grumble that citation analysis should be changed to account for this.

GLOBAL REACH

One important function of the year of chemistry will be to bring to light hidden contributions to science and society at large. Certainly, there is no shortage of global problems to which chemists can contribute solutions, such as the search for clean energy. And the Royal Society of Chemistry in London claims that some 20% of Britain's gross domestic product is already down to the work of chemists.

Chemistry is a mature field, but its exciting, productive and influential days are far from over. In the past few years alone, *Nature* has published cutting-edge research from chemists across the discipline. Some have probed the properties of the emerging material graphene (see page 14) and experimented with the new tool of DNA nanotechnology to design and assemble molecular machinery. Others have published details of new and improved materials for energy-storage devices. Organic chemists have shown how complex molecules can be made without the need for the sheltering protective groups commonly used to shield fragile molecules from decomposition during reactions. And the use of gold nanoparticles in oxidation reactions has revealed fundamental details of catalysis — the engine of much of chemistry and of life — and bolstered the wider work of chemists who strive to develop cleaner and less polluting industrial processes (see page 18).

If chemistry can truly be both good and bad, as Sandler observed, then much the same can be said for all the sciences, as for all human endeavour. But when chemistry is good, it is very, very good. It deserves its celebration. ∎

Legal highs: the dark side of medicinal chemistry

Synthetic chemist **David Nichols** *describes how his research on psychedelic compounds has been abused — with fatal consequences.*

This is the start of the international year of chemistry, intended to celebrate the contribution of my field to mankind's well-being. Yet, during the previous year it has become disturbingly clear to me that some of my scientific contributions may not be aiding people's well-being at all. In fact, they could be causing real harm.

A few weeks ago, a colleague sent me a link to an article in the *Wall Street Journal*. It described a "laboratory-adept European entrepreneur" and his chief chemist, who were mining the scientific literature to find ideas for new designer drugs — dubbed legal highs. I was particularly disturbed to see my name in the article, and that I had "been especially valuable" to their cause. I subsequently received e-mails saying I should stop my research, and that I was an embarrassment to my university.

I have never considered my research to be dangerous, and in fact hoped one day to develop medicines to help people. I have worked for nearly four decades synthesizing and studying drugs that might improve the human condition. One type is designed to alleviate the symptoms of Parkinson's disease, and it works superbly in monkey models of the disease. That same research seeks drugs to improve memory and cognition in patients who have schizophrenia, one of the most devastating human conditions. The other substances I work on are psychedelic agents such as LSD and mescaline. It's in that latter area of research that I have published papers about numerous molecules that probably have psychoactive properties in humans. It seems that many of these are now being manufactured and sold as 'legal highs'.

I first became aware that unknown amateur chemists were watching my papers more than a decade ago. My laboratory was doing research on 3,4-methylenedioxymethamphetamine (MDMA or ecstasy), a project we had started in 1982, before most people had even heard of the drug. We wanted to discover how MDMA worked in the brain because we thought drugs like it might help in psychotherapy. In the process, we studied many molecules that had structures similar to MDMA. One was 4-methylthioamphetamine, or MTA, which could inhibit the enzyme that breaks down serotonin in the body. Between 1992 and 1997, we published three papers on the effects of MTA in rats, including a study showing that MTA might have potential in the treatment of depression, and could possibly be superior to currently marketed drugs.

Without my knowledge, MTA was synthesized by others and made into tablets called, appropriately enough, 'flatliners'. Some people who took them died. Now, any knowledgeable person who had carefully read our papers might have realized the danger of ingesting MTA. It not only caused the release of serotonin from neurons, but also prevented the breakdown of this neurotransmitter, potentially leading to a dangerous serotonin syndrome that can sometimes prove fatal. My laboratory had shown that rats perceived the effects of MTA as being like those of ecstasy. It seemed that that was the sole motivation for its illicit production and distribution to humans. I was stunned by this revelation, and it left me with a hollow and depressed feeling for some time. By 2002, six deaths had been associated with the use of MTA. It did not help that I knew some of these fatalities were associated with the use of multiple drugs, or had involved very large doses of MTA. I had published information that ultimately led to human death.

There really is no way to change the way we publish things, although in one case we did decide not to study or publish on a molecule we knew to be very toxic. I guess you could call that self-censure. Although some of my results have been, shall we say, abused, one cannot know where research ultimately will lead. I strive to find positive things, and when my research is used for negative ends it upsets me.

Over the past year or so, I have begun to get more e-mails asking questions about the human effects of other materials that my laboratory had studied. Forensic laboratories began to send me requests for samples of drugs that they suspected were appearing on the black market, but which were so new that there are no analytical standards. Thankfully, most of the other molecules we have published on could not kill, at least not at reasonable dosages. But at very high doses, or mixed with other substances, they could become part of a lethal mix.

We never test the safety of the molecules we study, because that is not a concern for us. So it really disturbs me that 'laboratory-adept European entrepreneurs' and their ilk appear to have so little regard for human safety and human life that the scant information we publish is used by them to push ahead and market a product designed for human consumption. Although the testing procedure for 'safety' that these people use apparently determines only whether the substance will immediately kill them, there are many different types of toxicity, not all of which are readily detectable. For example, what if a substance that seems innocuous is marketed and becomes wildly popular on the dance scene, but then millions of users develop an unusual type of kidney damage that proves irreversible and difficult to treat, or even life-threatening or fatal? That would be a disaster of immense proportions. This question, which was never part of my research focus, now haunts me. ∎

> ## I WAS STUNNED. I HAD PUBLISHED INFORMATION THAT ULTIMATELY LED TO HUMAN DEATH.

David Nichols *is the Robert C. and Charlotte P. Anderson Distinguished Chair of Pharmacology at Purdue University in West Lafayette, Indiana.*
e-mail: drdave@purdue.edu

SEVEN DAYS *The news in brief*

Solar subsidy cuts

Spain's government is slashing subsidies to its photovoltaics industry by around 30%, or €740 million (US$970 million) per year, as part of plans to save more than €4.6 billion in electricity costs over the next three years. The measures, announced on 23 December, cut generous feed-in tariffs (the price an electricity utility must pay to generators of solar energy) by capping the subsidized hours that solar plants can sell. Solar industry investors protested that the changes would damage the industry.

Rare-earth quotas

Tensions over trade in rare-earth elements rose after China's ministry of commerce cut export quotas. On 28 December, it announced that the first batch of rare-earth mineral exports for 2011 would total 14,446 tonnes, 35% below the allowed quota in the first half of 2010 — and said that it would levy new export duties on rare-earth elements and alloys. Rare-earth elements are quite plentiful, but China controls more than 90% of their supply. They are used as catalysts and in technology including magnets, car batteries, wind turbines and mobile phones.

Plant catalogue

A working list of known land plant species — a key tool for conservation efforts — was released at the end of 2010 (see www.theplantlist.org). This eases a major problem for researchers: it is impossible to find everything published about a plant unless you know all of its scientific synonyms. The screwpine *Pandanus tectorius* (**pictured**), for example, has 294 synonyms. Researchers at the Royal Botanic Gardens, Kew, UK, and the Missouri Botanical Garden in St Louis sorted through 1.24 million plant names from a number of data sets, whittling them down to about 300,000 separate species linked to almost 480,000 synonyms.

Emissions control

The US Environmental Protection Agency's first controls on greenhouse-gas emissions came into effect on 2 January. For now they only affect the fuel efficiency of new vehicles and the issuing of permits for new industrial emitters of large amounts of carbon dioxide. But the agency plans to introduce tougher regulations over the next few years. Its authority to regulate greenhouse-gas pollution (under the Clean Air Act) is being challenged by opposition in Congress and by litigation.

Brazil joins Europe

Brazil has become the first country outside Europe to become a member state of the European Southern Observatory (ESO), after signing a formal accession agreement on 29 December. The deal — subject to parliamentary ratification — gives Brazilian astronomers access to telescopes at three sites in Chile operated by ESO, which is headquartered in Garching, Germany.
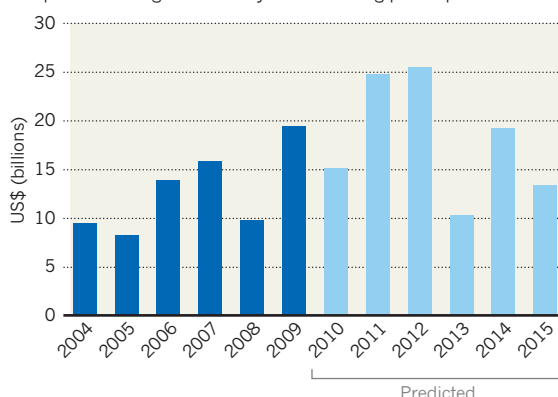
Nuclear treaty

The US Senate on 22 December ratified a deal with Russia to reduce nuclear arms, paving the way for new bilateral inspections. The New Strategic Arms Reduction Treaty (New START) was signed by the US and Russian presidents in April 2010, and requires the two nations to keep the numbers of their nuclear warheads and launch vehicles below limits that are lower than those agreed in a previous treaty; it also allows hands-on inspections to verify the numbers. The treaty awaits ratification by Russia's parliament. See page 9 for more on science and technology bills in the US Congress.

BUSINESS WATCH

In 2011 and 2012, the drug industry will face the worst effects of its 'patent cliff', as several major drugs face competition from generics. The world's best-selling drug — Pfizer's cholesterol-lowering medication Lipitor (atorvastatin) — may face competition from generics maker Ranbaxy of Gurgaon, India. Other blockbusters due to come off patent include Sanofi-aventis's anti-clotting drug Plavix (clopidogrel) and Eli Lilly's schizophrenia drug Zyprexa (olanzapine).

EDGE OF THE PATENT CLIFF

Blockbuster drugs with combined US sales of $25 billion face competition from generics this year after losing patent protection.

US$ (billions)

Year	US$ (billions)
2004	9.5
2005	8.5
2006	14
2007	16
2008	9.8
2009	19.5
2010	15
2011	24.8
2012	25.7
2013	10.5
2014	19.3
2015	13.5

Predicted

NEWS IN FOCUS

Cold comfort: the changing landscape of the US Congress could curtail the Democrats' plans for science.

R. T. NOWITZ/CORBIS

POLICY

US science faces big chill

Spending cuts and political battles loom on the horizon.

BY IVAN SEMENIUK, MEREDITH WADMAN, EUGENIE SAMUEL REICH AND JEFF TOLLEFSON

"We came here to do a job. We got much of it done." So said Nancy Pelosi, former speaker of the US House of Representatives, at a press briefing on 22 December 2010. For US researchers, the occasion was significant. Pelosi, standing alongside Bart Gordon, outgoing chairman of the science and technology committee (see page 10), had just signed a bill reauthorizing America COMPETES, the act that includes funding increases for basic research and educational programmes via the National Science Foundation (NSF) and other agencies.

For Pelosi and House Democrats there was reason to be upbeat. Despite losing control of the House in November's midterm elections, the final session of Congress was unexpectedly productive, including legislation that will aid science and technology. In addition to America COMPETES, it passed a bill giving the Food and Drug Administration sweeping new powers to regulate food safety. And the Senate ratified New START, a bilateral treaty with Russia to reduce nuclear arms.

Yet the last-minute push stands in contrast to a record that in many ways fell short of the expectations raised when the Democrats took charge of Congress in 2006, and Barack Obama became president in 2008. A House bill that included a cap-and-trade scheme for carbon emissions died in the Senate last year. Another to explicitly permit federal funding of human embryonic stem-cell research never made it to a vote. Legislation to tighten controls on toxic chemicals was introduced but failed to gain momentum.

In the Congress that convenes this week, the House of Representatives will be in Republican hands, making it even more likely that politically volatile bills, such as those involving climate change, environmental regulations and stem cells, will fail to advance. Less certain is the extent to which science will be hit by the spending cuts promised by the Republicans.

Currently, most of the US government is in a fiscal holding pattern, kept afloat by a 'continuing resolution' that extends the 2010 budget until 4 March. For many agencies this is already bad news. At the $31-billion National Institutes of Health (NIH), for example, it means a hoped-for increase of $750 million in the 2011 budget has effectively evaporated.

The best possible outcome for the NIH in 2011 would be "to maintain funding at 2010 levels", says David Moore, senior director for governmental relations at the Association of American Medical Colleges in Washington DC. But as the Republicans have proposed reducing non-defence spending to 2008 levels, even that seems a remote prospect. "It is relatively easy to see some sort of across-the-board cut imposed on federal agencies," says Moore.

NASA is in a particularly tricky position. In the NASA Authorization Act of 2010, which sets out the agency's budget request and programme priorities for the coming year, Congress voted for an additional space-shuttle flight in 2011, a new programme to develop successor vehicles and a halt to Constellation, President George W. Bush's plan to send astronauts ▶

back to the Moon. Yet under the continuing resolution, the agency must keep spending on the Constellation programme, which is budgeted at around $7 billion per year. Keith Cowing, editor of NASAWatch.com, says it is hard to see how NASA can finance the shuttle flight while juggling everything else. "It's like trying to take a large truck and do a sudden left turn," he says. NASA may be especially susceptible to political wrangling in the new Congress because many influential Republicans, including Gordon's successor on the science committee, Ralph Hall (Texas), have NASA centres in their districts or states and support a strong manned-spaceflight programme. Their resistance will make it harder for Obama to give the agency a fresh direction.

Hall spoke against the reauthorization of America COMPETES, arguing that he would rather scrutinize and vote on each science programme funded by the bill than give agencies such as the NSF a wholesale increase. He has already singled out for criticism the $900-million allocated to ARPA-E, a new agency of the US Department of Energy that promotes advanced energy research. ARPA-E originally had bipartisan support and Hall's scepticism came as a disappointment to Charles Vest, president of the National Academy of Engineering, who has championed COMPETES. "I believe the uniqueness and full importance of ARPA-E was missed," Vest says.

A battle over energy policy may well be inevitable because several incoming Republicans have expressed scepticism about climate change. Incoming chairman of the House Committee on Energy and Commerce, Fred Upton (Republican, Michigan), represents a moderate voice, but he has appointed more conservative lawmakers to chair his subcommittees.

John Shimkus (Republican, Illinois), who heads the new environment and economy subcommittee, made headlines in November when he quoted scripture to make the case that God would not destroy Earth through global warming. Ed Whitfield (Republican, Kentucky), who will head the new energy and power subcommittee, has already attacked the Obama administration's regulatory initiatives, in particular suggesting that the Environmental Protection Agency has not provided "compelling scientific evidence" to justify its climate regulations.

With government scaling back spending, and concerns about economic growth at the forefront, any attempt to curtail emissions is likely to face strong opposition on economic grounds. Abraham Lincoln famously observed that "a house divided" cannot stand. For US scientists, 2011 may be the year that demonstrates that a government divided cannot move. ∎

C. OWEN/NATURE

Q&A Bart Gordon
Going out on a high

After 25 years in the US Congress, Bart Gordon (Democrat, Tennessee), the chairman of the House Committee on Science and Technology, officially stepped down on 3 January. An ardent champion of science, he has served on the committee throughout most of his time in Washington DC and has led it since 2007. In 2007 he was instrumental in creating the America COMPETES Act, which aims to double funding for basic research in the physical sciences over a ten-year period and improve science education in US schools. One of his final acts as chairman was successfully shepherding a renewal of COMPETES through a politically polarized Congress. The act will be signed into law by President Barack Obama this week.

Why is America COMPETES so important to you?
The United States needs to be able to increase its investment in research. Research leads to innovation; innovation leads to jobs; jobs leads to more taxes, which then pay for more research. I have a 9-year-old daughter and I am concerned about her future and our country's future. There are approximately 7 billion people in the world, and half of those who are working make less than $2 a day. So unless we continue to innovate and unless we have a skilled workforce, we are going to see our standard of living decrease.

How will it be possible to invest more in research when many are calling for cutbacks?
It's going to be a challenge. We're seeing a little increase in the public-sector research dollars and we're seeing a decrease in private-sector funding. In the rest of the world many are

trying to do both: their private and public sectors are investing more. We're going to have to rally the private sector, the universities and everyone who cares about this to show its importance.

Is science playing a greater or lesser part in US policy-making?
I think President Obama has put a strong marker down that he wants to see science take a greater role. He has brought together an unheard of number of high-calibre scientists [in his administration] that I think are helping to do this. The thing that I've found is that whether it's John Holdren [director of the White House Office of Science and Technology Policy] or [energy secretary] Steve Chu, you have these top-notch scientists in their own fields

that, prior to coming to Washington, have come to know each other, and so it really gives them the ability to talk outside their bureaucratic silos.

Are there examples of where science can lead to better policy?

I think the business community is very interested in seeing science play a larger part in policy-making. Take nanotechnology and synthetic biology, for example. They are going to be big growth industries for our future. But there are health and safety concerns — and there can either be a perceived risk to health and safety or an actual one. So it is important that we get an early start on transparent research in health and safety in these two areas so that the public can feel comfortable that if there is a problem, then we're getting it off the shelf, and if there isn't, then there's a body of evidence that they can see.

Are you disappointed that Congress has not passed a climate bill?

I voted for the House bill but I'm less interested in a particular bill than I am in trying to deal with the problem. I'm very concerned that we're outliers compared with the rest of the world. In term of our international stature, that really has hurt, particularly in Europe.

You are known for reaching out to European legislators. What has been your aim in doing so?

Between the European Parliament and the US Congress, you have two legislative bodies that represent 800 million people and over half the world's GDP. The European Union and the United States also share similar cultures and wage scales. I really think it's important for us to be collaborating, intellectually and financially, in bringing more symmetry to standards and regulations and reducing barriers between our two regions. We have a common interest, for example, in clean energy. Some of those areas, such as next-generation nuclear power, carbon capture and sequestration, and fusion are hugely expensive. I think those are areas where we can collaborate.

Do you have any advice for your successor as chairman?

Try to maintain the civility that allowed us to work together. I tried to bring the Republicans in early to make them a part of the process. It made our bills better, and because of that we were able to go to the floor with a unified effort and pass legislation in a bipartisan manner — and if you want legislation to continue here, it needs to be bipartisan. ■

INTERVIEW BY IVAN SEMENIUK

Chávez squeezes scientific freedom

A string of new laws and a presidential power grab unsettle researchers in Venezuela.

BY ANNA PETHERICK

Venezuela's beleaguered scientists are facing renewed pressure from their government, which this week assumes control of levies from private companies that represent one of the main sources of research funding in the country. Meanwhile, President Hugo Chávez has gained fresh powers to enact legislation by decree, which some researchers fear he will use to close universities or curtail academic freedom.

These changes were hurried through Venezuela's National Assembly following the national elections on 26 September — when Chávez's party lost its two-thirds majority. The new assembly convenes on 5 January. "There are problems particularly for us in science," says mycologist Gioconda San-Blas, an emeritus professor at the Venezuelan Institute for Scientific Research in Caracas. "First we have a new law for science and technology, then restrictions to the Internet. Now there is a new law relating to universities as well."

The changes to the science and technology law, known as LOCTI, are manifold. Enacted in 2005, LOCTI provided a boost to science funding in Venezuela by requiring larger companies to plough money into research — which could be done either in-house, or at a university or research institute chosen by the company. Today, LOCTI funds amount to 3–4% of Venezuela's gross domestic product (GDP), compared with government science funding of about 0.5% GDP. Although LOCTI funds did not always reach the best public labs, some companies "gave generously to university projects with very good results", says Jaime

Requena, a former president of the Institute for Advanced Studies in Caracas, who was dismissed in 2009 after criticizing the government. But "the new version forces all private enterprises to surrender their LOCTI contribution to an office within the Ministry of Science and Technology", he says. "Now the destiny of all collected funds from private sources will be decided by the government according to the 'Socialist Plan for the Nation.'"

The amendments also narrow the fields of enquiry that can receive LOCTI funds to just four categories — climate change, energy innovation, building materials and urban development — and enable almost anyone to carry out the research, regardless of their qualifications. "According to the government, everyone can do science," says San-Blas, who worries that science in the country will become less professional as a result. The changes were approved without consultation with the research community.

Other legal changes mean that university budgets will now be controlled by 'communal councils' made up of local citizens, which will also elect university vice-chancellors. A new telecommunications bill mandates Internet providers to censor web pages according to government guidelines, potentially restricting scientists' access to information.

Orlando Albornoz, a sociologist at the Central University of Venezuela, Caracas, fears that Chávez's augmented presidential powers could be used to close universities that host professors who vocally oppose him. "If he closes down the autonomous universities he may face an ugly fight, but having the power in his hands he will oblige his enemies to negotiate on his terms," he says. "I see more and more control." ■

MORE ONLINE

2010 IN REVIEW

Nature looks back at a dramatic year in science **www.nature.com/2010**

FOLLOW *NATURE*'S NEWS

● On Twitter **twitter.com/naturenews**
● On FriendFeed **friendfeed.com/naturenews**
● Receive daily or weekly e-mail alerts **go.nature.com/jdlhcz**
● Subscribe to the RSS feed **go.nature.com/i2npjl**

PROSPECTS

New year, new science

Nature looks at key findings and events that could emerge from the research world in 2011.

THE EEMIAN REVEALED

The North Greenland Eemian Ice Drilling (NEEM) project reached bedrock in July 2010, at a depth of more than 2,500 metres. The fruits of that effort should soon be seen, now that researchers are analysing gas and particles trapped inside the ice core to reveal details of the climate of the Eemian inter-glacial period (130,000–115,000 years ago), when the average global temperature was about 5 °C warmer than today.

GWAS PROVE THEIR WORTH

Genome-wide association studies (GWAS) have uncovered plenty of links between diseases and particular regions of the genome, but frustratingly haven't revealed much about the biochemistry behind these associations. In 2011, expect to see real mechanistic insights explaining how genes, and non-coding regions, affect the medical conditions they have been linked with. Metabolism, obesity and diabetes are among the hottest targets.

STEM CELLS: READY FOR STUDY

Researchers have learned how to reprogram people's cells into induced pluripotent stem (iPS) cells, and on from that into other cell types: skin cells can be turned into nerve cells, for example. Patient-derived iPS cells will increasingly be used as models for studying medical conditions — particularly those, such as psychiatric disorders, for which there are no good animal models, and little understanding of what is happening inside cells. They will also be used to screen potential drugs, and to probe why existing drugs help some patients but not others.

GENOME-SEQUENCING EXPLOSION

This year should surely see the price of human-genome sequencing dropping to US$1,000 per genome. As next-generation sequencing machines reach the market, the number of fully sequenced human genomes will skyrocket.

THAT DAMNED ELUSIVE HIGGS

Although it is unlikely that the Higgs boson will be spotted this year by the Large Hadron Collider near Geneva, Switzerland, there's a good chance the collider will turn up something,

such as evidence for supersymmetry — the theory that every known fundamental particle has an undiscovered, superheavy partner. Meanwhile, Fermilab's Tevatron in Batavia, Illinois, is pushing for an extension beyond its September 2011 shutdown, and still hopes to hit the Higgs jackpot.

DARK MATTER'S MOMENT OF TRUTH

A number of underground experiments, such as XENON100 at Italy's Gran Sasso National Laboratory near L'Aquila, and the Cryogenic Dark Matter Search (CDMSII) in northern Minnesota's Soudan Mine, are hunting for dark matter particles and expect to release results in 2011.

HEPATITIS C TREATMENT

Eagerly anticipated drug approvals in 2011 include a decision by the US Food and Drug Administration on telaprevir, which

Will the National Ignition Facility ignite 2011?

could provide relief for the 3% of the world's population infected with the hepatitis C virus. The drug was developed by Vertex Pharmaceuticals in Cambridge, Massachusetts.

ANOTHER EARTH

Planet-hunters anticipate that NASA's Kepler telescope will reveal an Earth-like planet orbiting a Sun-like star. It has already spotted hundreds of planets outside the Solar System, although full data have not yet been released.

SYNTHETIC BIOLOGY: THINK MULTICELLULAR

No longer will scientists have to cram complicated synthetic biology into a single cell. Last year, researchers engineered an entire colony of bacteria to periodically fluoresce in unison, and we can expect many more papers exploring the behaviour of collections of cells. The goal is to exploit this teamwork to give bacteria useful functions such as producing medicinal drugs.
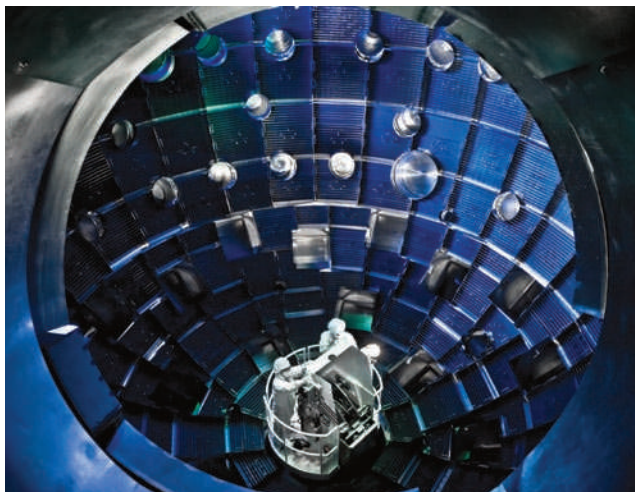
LAST OF THE SHUTTLES

The final flight of NASA's space-shuttle fleet is scheduled for April, when it will deliver the Alpha Magnetic Spectrometer (AMS) to the International Space Station to search for antimatter and dark matter. However, the US Congress may authorize another shuttle outing in November. If the second test launch of Dragon, the craft developed by commercial spaceflight firm SpaceX in Hawthorne, California, proves successful, the launch of a private spacecraft with crew or cargo is not out of the question.

SOLAR-SYSTEM EXPLORERS

In March, NASA's Messenger mission is due to become the first craft ever to orbit Mercury, and the agency's Dawn probe will orbit one of the biggest members of the asteroid belt, Vesta, in August. Other planned space launches include Juno, which will orbit Jupiter's poles; the GRAIL mission, twin spacecraft due to measure the Moon's gravitational field; and the Mars Science Laboratory, a car-sized rover that will explore the red planet.

SUPERLASER FLIRTS WITH FUSION

California's National Ignition Facility (**pictured**), the world's most powerful laser, is inching its way to triggering ignition, when fusion reactions in a target of hydrogen isotopes should produce more energy than the laser delivers. Experts give even odds that the laser, at the Lawrence Livermore National Laboratory, will succeed this year.

PROBING HOME

The European Space Agency's satellite GOCE, which is designed to measure Earth's gravity field in unprecedented detail, will publish results next year that will be used to help monitor sea level rise. Meanwhile, NASA's Aquarius satellite will launch to measure ocean salinity, and Glory will monitor solar irradiance and aerosols. ∎

COMPILED BY RICHARD VAN NOORDEN WITH HEIDI LEDFORD AND ADAM MANN

One of the detectors of the neutrino 'telescope' is lowered into place.

ASTROPHYSICS

IceCube completed

A neutrino observatory encased in ice at the South Pole could spot new physics.

BY ADAM MANN

On the day that researchers lowered the final detector of the IceCube Neutrino Observatory into a 2.5-kilometre-deep hole, the Antarctic sun was nearly as high as it gets and the temperature a balmy −23 °C. "It is quite warm," reported team member Albrecht Karle, a physicist at the University of Wisconsin in Madison. Just a week before the 18 December event, he adds, temperatures had averaged about 10 °C colder.

Karle and his collaborators aren't complaining. The deep Antarctic cold has created an ideal natural medium for detecting high-energy neutrinos. At a depth of 1.5 kilometres below the surface, the sheer weight of the overlying layers at the South Pole keeps the Antarctic ice sheet free of air bubbles and thus perfectly clear. Within its dark, transparent depths, even a faint flash of light can be spotted at some distance — including the kind of flash signalling that a fast-moving neutrino has hit an oxygen atom sitting in the ice and produced a muon.

The detector, a basketball-sized optical sensor, covered in the collaborators' signatures, was the 5,160th to be placed since construction began on IceCube in 2005. Since then, every Antarctic summer, researchers have used a jet of near-boiling water to drill holes in the ice. At every hole, a kilometre-long string of detectors was lowered down and the hole allowed to refreeze.

Now complete, at a cost of about US$271 million, IceCube monitors a cubic kilometre of ice — the size required for the experiment to have a realistic chance of spotting its rare and elusive quarry. But as the instrument has grown, so has its scope. Over the years, the team has realized that IceCube might also shed light on a broader range of questions, including an understanding of physics beyond the standard model.

"The most interesting things we're doing now, we didn't think of when it was proposed," says Francis Halzen, a physicist at the University of Wisconsin in Madison and the IceCube collaboration's principal investigator.

After photons, neutrinos are the most common particle in the Universe — trillions stream

The final detector: signed, sealed and delivered.

through your body every second. They are created in the nuclear-reactor centres of stars such as the Sun and by a variety of other objects in space. And because they are electrically neutral, unaffected by magnetic fields that permeate space, their straight-line trajectories point back to their sources.

If high-energy neutrinos turn up in IceCube's data as expected, they will provide astronomers with information they could not get from photons, says Eli Waxman, an astrophysicist at the Weizmann Institute of Science in Rehovot, Israel. One mystery it could crack is the source of ultra-high-energy cosmic rays (UHECRs), a poorly understood class of particles that carry billions of times the energy achievable in today's most powerful particle accelerators.

The two candidates currently suspected of generating UHECRs — active galactic nuclei and γ-ray bursts — are both powered by the accretion of mass onto a black hole. Yet the process by which these objects convert gravitational energy to radiation and particles is one of the fundamental open questions in astrophysics, says Waxman. IceCube could simultaneously solve the UHECR puzzle and provide qualitatively new information on the way matter behaves under strong gravity.

IceCube's size also makes it ideal for understanding completely new physical phenomena, says Bill Louis, a physicist at Los Alamos National Laboratory in New Mexico and co-spokesman for the neutrino experiment MiniBooNE at Fermilab in Batavia, Illinois. As neutrinos travel, they may oscillate between three known types — electron, muon and tau — but results published last October suggest that MiniBooNE may have observed muon antineutrinos turning into a fourth type, called a sterile neutrino (A. A. Aguilar-Arevalo *et al. Phys. Rev. Lett.* **105,** 181801; 2010). The energy range over which IceCube searches makes it perfectly configured to discern this fluctuation, says Louis, which means that it might be able to confirm or overturn the strange MiniBooNE result, he adds.

IceCube might also help to reveal the nature of dark matter, an invisible component of the Universe that may consist of one or more types of weakly interacting particles (WIMPs). In some models, WIMPs are predicted to accumulate at the Sun's core. Such a concentration would increase the odds of dark-matter particles colliding with one another and annihilating to produce high-energy neutrinos, says Robert Stokstad, a physicist and project director for IceCube at Lawrence Berkeley National Laboratory in California. If IceCube could find these neutrinos and confirm that they originate in the Sun, it would be the first direct detection of dark matter, he says.

Waxman says that the most exciting possibility is that IceCube will detect something entirely unexpected. "In the most basic respect," Waxman says, "IceCube is providing us with a new way to look at the Universe." ∎

B. GUDBJARTSSON/NSF

LLNL
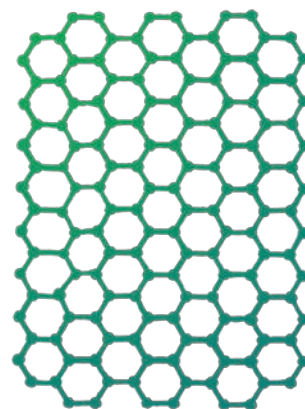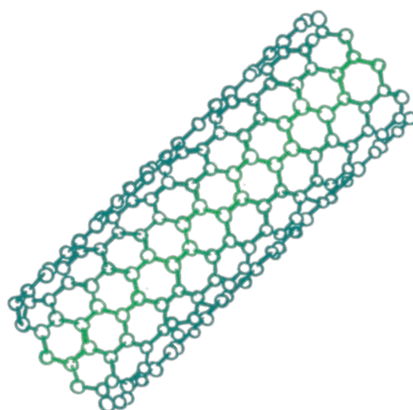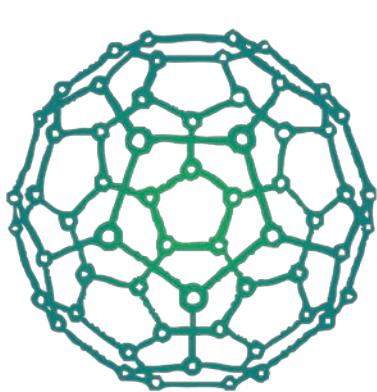
THE TRIALS OF NEW CARBON

Researchers have spent 25 years exploring the remarkable properties of fullerenes, carbon nanotubes and graphene. But commercializing them is neither quick nor easy.

By Richard Van Noorden

In fairy tales, third place is often the best: it's usually the third casket that contains the treasure, and the third child who finds fame and fortune. And so it may be for graphene, the third and most recently discovered form of 'new carbon'. The football-shaped fullerenes[1], discovered in 1985, and the hollow cylindrical carbon nanotubes[2], first characterized in 1991, have so far had a limited impact on industry. But now graphene, a one-atom-thick flat sheet of carbon, seems to be surrounded by favourable omens — not the least of which is the speed with which groundbreaking experiments on its properties were rewarded with the 2010 Nobel Prize in Physics.

It has been just six years since Nobel laureates Andre Geim and Kostya Novoselov at the University of Manchester, UK, first reported using sticky tape to peel atomically thin layers of graphene from lumps of graphite[3]. But the material — essentially just an unrolled nanotube — has turned out to have properties just shy of miraculous: a single layer of graphene is simultaneously the world's thinnest, strongest and stiffest material, as well as being an excellent conductor of both heat and electricity.

Graphene has been showered with media attention as companies vie to bring those attributes to market. Last year, graphene was the subject of around 3,000 research papers and more than 400 patent applications. South Korea is planning a US$300-million investment to commercialize the material, and companies ranging from IBM to Samsung are

testing graphene electronics — ultra-small, ultra-fast devices that might one day replace the silicon chip. The hype over graphene has reached such a pitch that a casual follower might wonder why it hasn't conquered the technological world already.

The reality is not such a fairy tale. Graphene's carbon forebears were once hyped in much the same way. Yet fullerenes have found hardly any practical applications. And although nanotubes have done better, they are costly to produce and difficult to control. Their subdued industrial impact is a lesson in just how hard commercialization of a new material can be.

Yet the story of nanotubes has some encouraging features. High-tech electronics applications are still years in the future, but a more low-tech application — nanotube-based conducting films for energy storage or touch screens — is much closer to commercialization. Another, comparatively straightforward use — nanotube-reinforced composite materials for aeroplanes and automobiles — is now reaching the market. Anticipating growing demand, nanotube manufacturers have scaled up production to many hundreds of tonnes a year.

For that very reason, the graphene manufacturers following in their wake may have hit on the right moment to start mass-producing the sheets. Graphene is being considered for the

same types of application as nanotubes, but it has some key advantages in ease of production and handling, and should benefit from two decades of research with nanotubes. That hindsight also means that graphene manufacturers have a better idea of which applications are worth chasing, and of how to avoid the false starts that nanotubes made in their first decade.

A CARBON PLAYGROUND

The remarkable properties shared by nanotubes and graphene arise from their common structure: an atomically thin mesh of carbon atoms arranged in a honeycomb pattern. Immensely strong carbon–carbon bonds produce an exceptionally high strength-to-weight ratio. Such is the strength of graphene, for example, that according to the Nobel prize committee, a hypothetical 1-metre-square hammock of perfect graphene could support a 4-kilogram cat. The hammock would weigh 0.77 milligrams — less than the weight of a cat's whisker — and would be virtually invisible.

The symmetry with which carbon atoms are arranged on the hexagonal lattice also allows both forms of nano-carbon to conduct electricity far more easily than the silicon used in computer chips. This means that they have much lower electrical resistance and generate much less heat — an increasingly useful property as chip manufacturers try to pack features ever more densely onto circuits.

Furthermore, even small variations in carbon structure can create a multitude of new properties. In graphene, for example, electronic
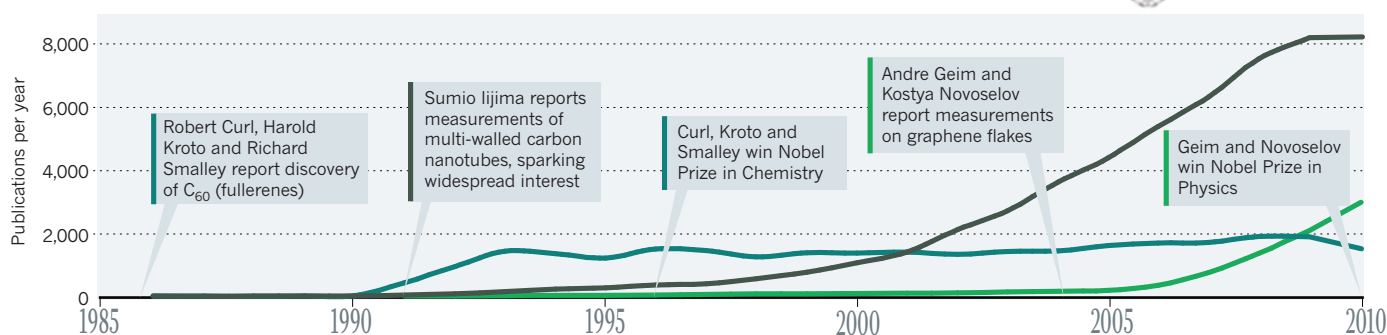
NEW CARBON IN THE LAB

Interest in graphene, indicated by publications per year, is accelerating faster than the buzz that surrounded carbon nanotubes in the 1990s.

— Carbon nanotubes
— Graphene
— Fullerenes

Publications per year

- Robert Curl, Harold Kroto and Richard Smalley report discovery of C_{60} (fullerenes)
- Sumio Iijima reports measurements of multi-walled carbon nanotubes, sparking widespread interest
- Curl, Kroto and Smalley win Nobel Prize in Chemistry
- Andre Geim and Kostya Novoselov report measurements on graphene flakes
- Geim and Novoselov win Nobel Prize in Physics

behaviour depends on the size of a given sheet, the presence or absence of defects in the sheet's lattice and whether it is lying on a conductive surface. In nanotubes, likewise, a given structure can be made semiconducting or metallic just by changing its diameter, length or 'twist' (the angle between the lines of hexagons and the direction of the tube). And there are differences between single tubes and those in which several cylinders are nested inside each other — called multi-walled nanotubes.

These properties have long sparked hopes of game-changing electronics applications. And researchers have made great progress — in the laboratory. In 1998, for example, physicists demonstrated a transistor made from a single, semiconducting nanotube[4]. And in 2007, researchers reported the synthesis of a carbon-nanotube-based transistor radio[5].

But for industrial-scale mass production of such circuits, the great variability of nanotubes is a curse. They are most commonly produced in a reactor, in which catalysts guide formation of the tubes from a carbon-rich vapour. This typically leaves a jumble of multi-walled, single-walled, semiconducting and metallic tubes of various lengths and diameters, all with different electronic properties. "Diversity is great until you have too diverse a population: then it becomes a real headache," says John Rogers, a physical chemist at the University of Illinois in Urbana-Champaign.

Only in the past five years have researchers worked out how to sort nanotubes into semiconducting and metallic types[6]. But there are further difficulties in assembling selected nanotubes in predetermined places on a chip and connecting these separate tubes together without compromising performance, so most physicists have come to believe that it is impractical for carbon nanotubes to replace silicon. "An integrated circuit would have to involve billions of identical carbon-nanotube transistors, all switching at exactly the same voltage," says Phaedon Avouris, who works on nanoscale electronics at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York. This

is not feasible with current technology.

Graphene offers a bit more cause for optimism. The highest-quality sheets are currently made by heating a wafer of silicon carbide in a vacuum, leaving a layer of pure graphene on the top surface. This method has fewer problems with uncontrollable variety from batch to batch than does nanotube synthesis, and the flat sheets that result are bigger and easier to handle than nanotubes.

But graphene has problems too. A single graphene sheet conducts charge so well that it is hard to make the current stop, something that must be solved if the material is ever going to be used in digital devices such as transistors, which control the flow of current like on–off switches. To change the material's electronic properties in the appropriate way — creating a 'band gap', or break in electron energy levels, which essentially turns it into a semiconductor — the sheet must be sliced up into thin

> ## "A lot of companies are bringing new capacity online at the same time right now."

ribbons. This is probably easier than trying to place billions of nanotubes on a chip, says Avouris — but it is still not currently possible with commercial technologies.

These processing difficulties suggest that graphene won't soon displace silicon chips. "There have been millions of person-years and trillions of dollars put into the development of silicon electronics," notes James Tour, an organic chemist who specializes in nanotechnology at Rice University in Houston, Texas. "Asking graphene to compete with silicon now is like asking a 10-year-old to be a concert pianist because we've been giving him piano lessons for the last six years."

In the meantime, nano-carbon structures may be more competitive in less demanding electronics, such as conductive flat films for

transparent electrodes in touch-screen displays or in solar cells. Bundles of dissimilar carbon nanotubes might very well provide enough conductivity for such electrodes, as might cheaper, lower-quality graphene sheets made by methods other than the silicon carbide process.

LOWERING THE SIGHTS

In June 2010, for example, a team led by Byung Hee Hong at Sungkyunkwan University in Suwon, South Korea, reported using carbon-rich vapour to deposit graphene films measuring 75 centimetres diagonally on copper plates, which are then etched away and recycled[7]. South Korean electronics giant Samsung is already testing this technique for use in commercial touch screens, which Hong estimates could be just two to three years away.
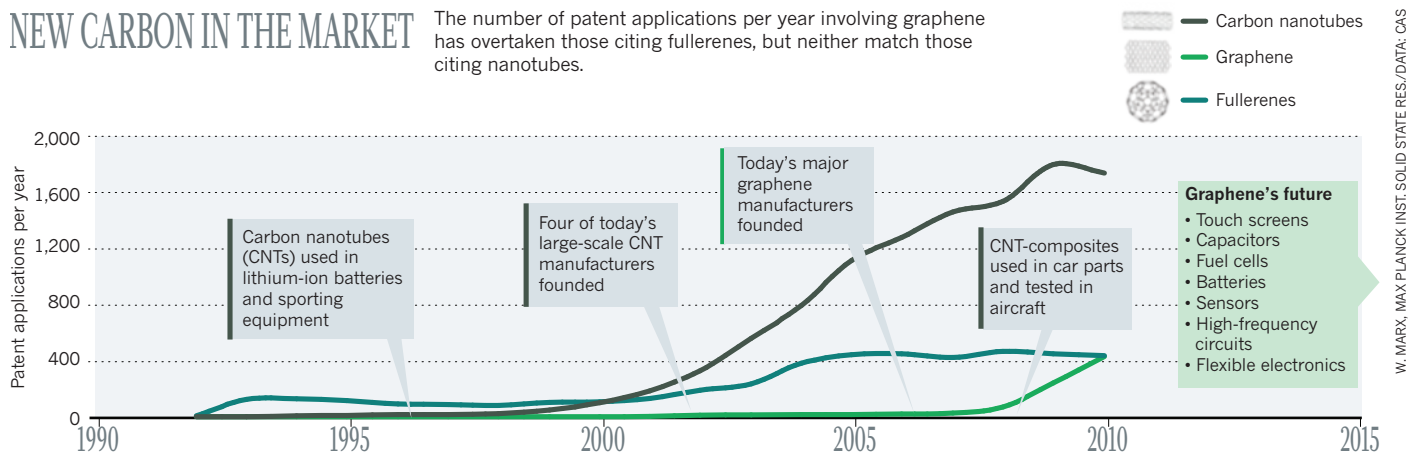
The question is whether the graphene films can compete with existing touch-screen materials such as indium tin oxide (ITO). Hong is optimistic; the cost of ITO has been increasing rapidly, because indium is scarce. But again, carbon nanotubes offer a cautionary tale. Early on, it was hoped that the tubes would form the television screens of the future, thanks to their ability to emit electrons from their tips to excite phosphors on the screen. In practice, competing plasma and liquid-crystal displays got better faster — and these are the screens most commonly used today.

One sweet spot for nano-carbon may be in the still-emerging market for flexible electronics. These are displays and sensors that could be worn on clothes, stuck to walls or printed on rollable sheets. Here, the only competition is from organic conducting polymers, because other materials cannot be printed on plastic. The performance of these polymers is quite low, says Rogers, so nanotubes and graphene circuits — which can be transferred to flexible substrates — could compete effectively.

But even these specialist electronics are still in the future. For now, the hundreds of tonnes of commercial nano-carbon being turned out every year are mostly going into composites for sporting goods, lithium-ion batteries and cars.

NEW CARBON IN THE MARKET

The number of patent applications per year involving graphene has overtaken those citing fullerenes, but neither match those citing nanotubes.

— Carbon nanotubes
— Graphene
— Fullerenes

W. MARX, MAX PLANCK INST. SOLID STATE RES./DATA: CAS

Patent applications per year (y-axis: 0, 400, 800, 1,200, 1,600, 2,000)

x-axis: 1990, 1995, 2000, 2005, 2010, 2015

Carbon nanotubes (CNTs) used in lithium-ion batteries and sporting equipment

Four of today's large-scale CNT manufacturers founded

Today's major graphene manufacturers founded

CNT-composites used in car parts and tested in aircraft

Graphene's future
• Touch screens
• Capacitors
• Fuel cells
• Batteries
• Sensors
• High-frequency circuits
• Flexible electronics

The aim is to disperse nano-carbon sheets or tubes within resins or polymers, so that they not only make the material tougher by blocking cracks that would otherwise spread, but also help to dissipate heat and electrical charge. For example, the Audi A4 car now has plastic fuel filters containing carbon nanotubes, which protect against static electricity. And nanotube additives in lithium-ion battery electrodes were one of the first nanotube applications marketed by Showa Denko, a Tokyo-based chemical-engineering company.

CUTTING COSTS

Basic processing problems with nanotubes initially hampered progress. They tended to clump together like tangled string as they came out of the reactor, making it hard to disperse the nanotubes evenly through plastic or resin. Despite improvements, this limits nanotube content to 1–2% by weight in the final product, compared with the 20–30% typical of conventional carbon fibre. The other problem was, and still is, cost. Materials such as steel, aluminium and plastics, and fillers such as carbon black, sell for just dollars or cents a kilogram, says David Hwang at Lux Research, a technology-evaluation company in New York. Meanwhile, multi-walled nanotubes retail for $100 a kilogram. The price is coming down as production scales up, but will drop only to about $50 a kilogram by 2020, according to Lux forecasts.

Composite-quality graphene has the potential to be a lot cheaper, although costs are currently similar. As Geim and Novoselov showed in 2004, graphene platelets of varying sizes can easily be peeled away from graphite[3], a raw material that costs a few dollars a kilogram. Graphene is also easier to disperse in a resin than are nanotubes.

But promising as this is, says Steve Hahn, a senior scientist with Dow Chemical's Ventures and Business Development Group in Midland, Michigan, the reality is that these applications are still niche. "I've been trying to find outlets for graphene for a couple of years," he says. But there is usually something quite a lot

cheaper that does the same job, says Hahn.

Michael Knox, president of XG Sciences, a start-up graphene-manufacturing company in East Lansing, Michigan, agrees. Adding graphene platelets to composites is not a transformational application, "it's an incremental improvement", he says. Yet that is not to be sniffed at. "If I could demonstrate a 10–20% improvement in a polypropylene composite at a reasonable price, I could probably sell a million tonnes of it a year — and car manufacturers would be pretty excited by that," says Knox.

The trick for young graphene-manufacturing companies is to find specific applications and then work out how to scale up production capacity without overstretching themselves. Vorbeck Materials in Jessup, Maryland, for example, has

> ## "Graphene will have its place, but it will just take longer than people think."

decided to focus on making graphene-based conducting inks. John Lettow, co-founder and president of Vorbeck, says that the inks will be in smart cards and radio-frequency identification tags in retail stores in the first quarter of 2011.

One near-term application may be supercapacitors, which use crumpled-up sheets of graphene to pack a massive surface area into a small space — to store more electric charge per gram than any other material. Other researchers are looking at using nano-carbon to make catalyst electrodes in fuel cells, or even to make water-purification membranes — but, as usual, finding clear advantages over existing materials such as activated carbon will be the problem.

Carbon nanotubes do have one property that graphene sheets don't: they can be very long. The nanotubes currently mixed into resins and plastics are typically short stubs, but Nanocomp Technologies in Concord, New Hampshire, says that it can spin long nanotube fibres into lightweight, electrically conducting yarns

or sheets that could replace copper wiring in some applications. "There are about 60 miles of copper wire in an aeroplane," says Nanocomp chief executive Peter Antoinette — so replacing this with the much lighter nanotube wire could save substantially on weight and fuel use.

Such activity is very encouraging for carbon nanotubes, says Hwang. "There was a huge amount of research that had to be done before nanotubes got to be commercially viable. Now if you look at the next five years, the commercial trajectory will be very different."

But have carbon nanotubes really taken a disproportionate time to get going? It takes 20 years or more for any new material to make an impact in industry, point out many nanotube makers. "Research on carbon fibre started in the 1950s; it took 15 years or so before aerospace and military used it — and we didn't hear about that until much later — and it wasn't until the mid-1970s that you started seeing commercial aircraft with small quantities of structures made out of carbon-fibre composites," says Brian Wardle, who directs the nano-engineered aerospace structures consortium at the Massachusetts Institute of Technology in Cambridge. Nanotubes may simply be on the usual trajectory from discovery to industry — and graphene may find that it follows the same path. "Graphene will have its place, but it will just take longer than people think," says Antoinette.

And what will happen in the meantime? "A lot of companies are bringing new capacity online at the same time right now," notes Hahn. "They will either go out of business or find a market somewhere. Whatever happens, it'll be a great lesson to all of us in how new materials are commercialized." ■

Richard Van Noorden is an assistant news editor at Nature.

1. Kroto, H. W. et al. Nature **318,** 162–163 (1985).
2. Iijima, S. Nature **354,** 56–58 (1991).
3. Novoselov, K. S. et al. Science **306,** 666–669 (2004).
4. Tans, S. J. et al. Nature **393,** 49–52 (1998).
5. Jensen, K. et al. Nano Lett. **7,** 3508–3511 (2007).
6. Arnold, M. S. et al. Nano Lett. **5,** 713–718 (2005).
7. Bae, S. et al. Nature Nanotech. **5,** 574–578 (2010).

IT'S NOT EASY BEING
GREEN

In the past two decades, the green-chemistry movement has helped industry become much cleaner. But mindsets change slowly, and the revolution still has a long way to go.

BY KATHARINE SANDERSON

By the latter half of the 1980s, the worldwide chemical industry knew that it had to clean up its act: its environmental reputation was dismal. Still fresh in the public mind was the 1984 disaster at Bhopal, India, where at least 3,000 people died and hundreds of thousands were injured by a toxic gas leak at a Union Carbide pesticide plant. Also fresh were memories of the 1978 Love Canal incident in Niagara Falls, New York, where the discovery of buried toxic waste forced the abandonment of an entire neighbourhood, and the discovery of dioxin contamination a few years later that forced the evacuation of an entire town — Times Beach, Missouri.

Even when companies did try to deal responsibly with their waste, which typically included volatile organic solvents and other hard-to-clean-up agents, the volumes were daunting. Global statistics were, and still are, fragmentary. But in the United States, according to the earliest systematic data gathered by the Environmental Protection Agency (EPA), some 278 million tonnes of hazardous waste were generated in 1991 at more than 24,000 sites. Not all of it came from chemical companies, but much of it did. More than 10% of the total, some 30 million tonnes, came from one firm alone: the Dow Chemical Company, headquartered in Midland, Michigan. And other firms, such as petrochemical giant Amoco, headquartered in Chicago, Illinois, and DuPont, of Wilmington, Delaware, were not far behind.

The result, as chemical companies struggled to deal with increasingly stringent environmental regulations, was an industry-wide move towards what is often called 'green chemistry' — a term introduced in 1991 by Paul Anastas, then a 28-year-old staff chemist with the EPA.

The goal of green chemistry was never just clean-up, explains Anastas, who is currently on leave from Yale University to head the EPA's research division. In his conception, green chemistry is about redesigning chemical processes from the ground up. It's about making industrial chemistry safer, cleaner and more energy-efficient throughout the product's life cycle, from synthesis to clean-up to disposal. It's about using renewable feedstocks wherever possible, carrying out reactions at ambient temperature and pressure — and above all, minimizing or eliminating toxic waste from the outset, instead of constantly paying to clean up messes after the fact. "It's more effective, it's more efficient, it's more elegant, it's simply better chemistry," says Anastas.

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

If the green-chemistry ideal is simple to state, however, achieving it has been anything but simple. Yes, says Eric Beckman, a chemical engineer at the University of Pittsburgh in Pennsylvania, "Companies these days are being very attentive to rendering their current processes greener." In 2009, for example, the total US output of hazardous waste was down by an order of magnitude over 1991, to 35 million tonnes. The largest generator in that year, DSM Chemicals in Augusta, Georgia, produced just 3.4 million tonnes.

But the greening of any given process is always a trade-off among benefits, feasibility and cost, says Beckman — and green is not always the winner. Furthermore, he says, industry's adoption of green chemistry has so far been focused mainly on incremental improvements in existing processes.

"It's embryonic at best," says Beckman, who speaks for many observers when he says that the real 'green revolution', in the form of processes redesigned from scratch and plants rebuilt from the ground up, is only just beginning.

The progress of green chemistry so far has been partly a matter of technical feasibility, as researchers have developed less toxic alternatives to conventional methods. A prime example is supercritical carbon dioxide: ordinary, non-toxic carbon dioxide that has been heated and pressurized above its 'critical point' of 31.1 °C and 7.39 megapascals, beyond which it behaves like both a gas and a liquid, and readily serves as a solvent for a wide range of organic and inorganic reactions. Other non-toxic replacements for solvents have been found among the ionic liquids: exotic cousins to ordinary table salt that happen to be liquid at or near room temperature.

THE E-FACTOR

Green chemistry's progress has also benefited from an awareness campaign by Anastas and his allies. A key first step was the 1991 coining of the name itself, says John Warner, president of the Warner-Babcock Institute for Green Chemistry in Wilmington, Massachusetts, who at the time was director of exploratory research at the Polaroid Corporation in Minnetonka, Minnesota. "Identifying green chemistry as a field of science differentiated it from a political and social movement," he says.

Another key step was the drawing up by Anastas and Warner of a set of principles intended to help scientists define and practise green chemistry (see 'The twelve principles of green chemistry'). And yet another came in 1995, when Anastas helped to persuade US President Bill Clinton to launch the Presidential Green Chemistry Challenge, which still awards five citations each year to companies and academics who have done an outstanding job of implementing the principles.

Mostly, however, green chemistry's progress has been a matter of corporate buy-in, as epitomized by its promotion by the chemical industry's own voluntary initiative, Responsible Care (www.responsiblecare.org), which works with national industry associations to improve the industry's health, safety and environmental performance. Founded in Canada in 1985, membership has grown from 6 national associations to 53.

The pharmaceutical sector has embraced green chemistry most enthusiastically, perhaps because it has the most to gain. Pharmaceutical plants typically generate 25 to 100 kilograms of waste per kilogram of product, a ratio known as the environmental factor, or 'E-factor'. So there is plenty of room to increase efficiency — and cut costs.

At drug-maker Pfizer, for example, the first laboratory synthesis of its anti-impotence drug sildenafil citrate (Viagra) had an E-factor of 105. But long before Viagra went on the market in 1998, a team at Pfizer's plant in Sandwich, UK, was rigorously re-examining every step of the synthesis. The researchers replaced all the chlorinated solvents with less toxic alternatives, and then introduced measures to recover and reuse these solvents. They eliminated the need to use hydrogen peroxide,

which can cause burns. They also eliminated any requirement for oxalyl chloride, a reagent that produces carbon monoxide in reactions and is therefore a safety concern. Eventually, Pfizer's researchers cut Viagra's E-factor to 8.

After that success, Peter Dunn, the leader of the Viagra synthesis team, became head of the more systematic green-chemistry drive started by Pfizer in 2001. Dunn says he is not free to talk about specific cash savings, but can point to sweeping changes made across the company. Pfizer has reduced the E-factor of the anticonvulsant pregabalin (Lyrica) from 86 to 9, he says, and has made similar improvements for the antidepressant sertraline and the non-steroidal anti-inflammatory celecoxib. "These three products alone have eliminated more than half a million metric tons of chemical waste," says Dunn.

CREATIVE CHEMISTRY

Nor is Pfizer alone; the pharmaceutical sector is so competitive that no company can afford to ignore green chemistry's potential savings. The Pharmaceutical Roundtable, first convened in 2005 by the American Chemical Society's Green Chemistry Institute, now has 14 member companies that jointly fund academic research in the field and share pre-competitive information.

In 2002, the chemicals giant BASF, based in Ludwigshafen, Germany, introduced an industrial-scale process that uses ambient-temperature ionic liquids to remove acid by-products from reaction mixtures — a common chemical manufacturing step that had previously been much more cumbersome. But BASF's embrace of green chemistry (which the company prefers to call 'sustainable chemistry') goes much further, notes Pete Licence, a green chemist at the University of Nottingham, UK. "You're getting sensible and joined-up thinking about the way that chemical plants are created," he says. "They have this integrated reaction system where the products and the by-products of reactions are actually the starting materials for the plant that is next door." The plants are also designed to maximize energy efficiency, Licence says: "Waste heat from one process is the warm-up for the feedstock for the next."

But the comprehensive restructuring required illustrates why the shift to green chemistry has been comparatively slow among bulk-chemicals manufacturers. These firms deal with products that are made in much larger volumes than pharmaceuticals, and their industrial processes are already highly optimized, with E-factors typically in the range of 1 to 5. Although it is possible to go much lower — E-factors for petrochemicals are on the 0.1 scale — doing so is not always economic. "Once you have a plant, it will run for 30 or 40 years because you have made a huge investment," says Walter Leitner at the Institute for Technical and Macromolecular Chemistry at the University of Aachen, Germany.

Nor does it always pay to be green in the speciality chemicals sector — as Thomas Swan and Company in Consett, UK, learned the hard way. In 2001, building on the work of chemist Martyn Poliakoff at Nottingham University, it opened the world's first continuous-flow reactor using supercritical carbon dioxide as a solvent. "It looked as if it could have been game-changing within the industry," says managing director Harry Swan. But when no government subsidies were forthcoming, the plant could not provide chemicals more cheaply than those made by the standard non-green methods, he says. So the facility was mothballed, and may soon be decommissioned and dismantled.

Other roadblocks to the adoption of green chemistry are technical. For example, even after decades of research, green solvents are not always more efficient than the widely used chlorinated solvents. Nor have chemists completely eliminated the need for catalysts containing precious or toxic metals — although Dunn, for one, is optimistic that this may eventually be possible through advances in enzyme technology. And how to make bulk chemicals from biomass and other renewable feedstocks, rather than from crude oil, is still an open challenge. "It's

IT'S MORE EFFECTIVE, IT'S MORE EFFICIENT, IT'S MORE ELEGANT, IT'S SIMPLY BETTER CHEMISTRY.

THE TWELVE PRINCIPLES OF GREEN CHEMISTRY

Paul Anastas and John Warner formulated these guidelines in the 1990s to provide a roadmap for chemists trying to implement green chemistry.

1 WASTE PREVENTION
Prevent waste from the start rather than treating or cleaning it up afterwards.

2 ATOM ECONOMY
Design synthetic methods to maximize the incorporation of intermediate materials into the final product.

3 SAFER SYNTHESES
Design synthetic methods to minimize the use and generation of toxic substances.

4 SAFER PRODUCTS
Design chemical products to carry out their function while minimizing their toxicity.

5 SAFER AUXILIARIES
Minimize the use of solvents and other auxiliary substances, and make them as innocuous as possible.

6 ENERGY EFFICIENCY
Minimize the energy used in chemical processes, and if possible, carry them out at ambient temperature and pressure.

7 RENEWABLE FEEDSTOCKS
Use biomass and other renewable raw materials whenever practicable.

8 DERIVATIVE REDUCTION
Minimize the potentially wasteful use of blocking groups and other temporary modifications of intermediates.

9 CATALYSIS
Prefer catalytic reagents — as selective as possible — over stoichiometric reagents.

10 DEGRADABILITY
Design chemical products for eventual disposal, so that they break down into innocuous compounds that do not persist in the environment.

11 POLLUTION PREVENTION
Develop methods for real-time monitoring and control of chemical processes that might form hazardous substances.

12 ACCIDENT PREVENTION
Choose processes and practices that minimize the potential for chemical accidents, including releases, explosions and fires.

---

a different way of looking at a chemical synthesis," says Leitner, who points out that the conventional problem gets turned on its head. Instead of starting out with a relatively simple hydrocarbon extracted from oil, and then adding side groups to the molecule to give it the desired properties, chemists have to start with the incredibly complex mixture of biomolecules typical of most renewable feedstocks, and get to what they want by snipping off pieces in a controlled manner.

But many advocates say that the most fundamental barrier to the wider adoption of green chemistry is mindset — which largely reflects the way chemists are taught. "In the United States, chemists get trained rigorously in chemistry, but don't see any engineering, product design, or life-cycle analysis," says Beckman. Or, as Anastas puts it, "you usually get the safety course that says, 'Wear your goggles and your coat, and don't blow things up — and by the way, here's the number to call if you do.' But I don't think that's the same as treating the consequences of what we do as an intrinsic part of our work".

This curricular conservatism may well reflect the often negative reactions of academic chemists to green chemistry. Especially in its early years, the field was seen as fuzzy and non-rigorous, recalls Neil Winterton of the University of Liverpool, UK, a former critic who has since become more accepting of the movement. The word 'green' conveyed the impression that certain techniques were being promoted for reasons of political correctness, he says. "It needed a little bit more fundamental underpinning to establish whether what was being proposed was or wasn't a major contribution to improving the efficiency of chemical processes."

Sceptics also questioned whether green chemistry was anything more than a trendy new buzzword used to get money for projects of dubious environmental value. "It's something that can dupe the public, it can dupe other scientists working in the area, and much, much more importantly, it can dupe decision-makers", concedes Licence.

Scepticism hasn't gone away entirely; a mention of green chemistry in a gathering of chemists can still provoke sighs and eye-rolling, says Warner. But scepticism has lessened as research has improved.

GREEN CHEMISTRY SHOULD JUST BE SECOND NATURE, THE DEFAULT VALUE.

The EPA, for example, has made notable progress in lifting one barrier to effective green chemistry, which is that researchers trying to create a new, non-toxic manufacturing process often don't know if a given compound is 'green' or not. No one has had the time or money to gather the toxicity data, which typically requires expensive animal testing.

The EPA's answer is a high-throughput screening project called ToxCast, which has been running at its Research Triangle Park facility in North Carolina since 2007. The ToxCast team has applied a battery of standard high-throughput biochemical assays, which measure such things as binding to cellular receptors, to 1,000 chemicals that already have animal toxicology data. These data have then been used to build statistical and computational models that attempt to predict any compound's toxicity from the assays alone.

A ToxCast prediction costs US$20,000 per chemical, compared with the $6 million to $12 million typical of animal toxicology tests, says Robert Kavlock, who oversees the project as head of the EPA's National Center for Computational Toxicology at Research Triangle Park. So if these models can be made reliable enough, he says, "then we've got a way to address the chemicals that we can't afford to test in animals" — and in the process, help companies to choose compounds that will make their chemistry truly green.

Now that Anastas is the EPA's research chief, he has been trying to spread the green-chemistry approach through staff meetings at the agency's labs across the country. He wants to move the EPA away from a culture of regulating and banning to one where products are designed to be synthesized in a way that reduces or eliminates the use of hazardous substances in the first place. As EPA chief Lisa Jackson puts it: "It's the difference between treating disease and pursuing wellness."

If that change in attitude happens, says Anastas, it will represent a "seismic shift" at the agency — "the culmination of the work of my career". But in a sense, he adds, it will also be just the beginning: "I believe that the ultimate goal for green chemistry is for the term to go away, because it is simply the way chemistry is always done. Green chemistry should just be second nature, the default value." ∎

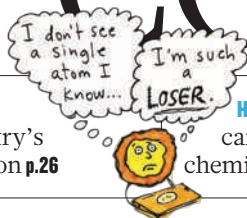**Katharine Sanderson** *is a writer based in Toulouse, France.*

COMMENT

Design innovations such as Princeton University's Frick Chemistry Lab need to be matched by a radical rethink of priorities and teaching methods.

Let's get practical

Chemistry needs an overhaul if it is to solve big global problems and advance fundamental understanding, say **George M. Whitesides** and **John Deutch**.

Chemistry is at the end of a century of expansion. In 1900, the chemical industry was in its infancy. The modern research university was still 50 years away, and the basic concepts of the field — the chemical bond, the laws of thermodynamics, theories of kinetics — were still being developed. In 2011, the industry is mature and fully embedded in society, and chemists have a good, semi-empirical grasp of many of the characteristics of molecules and reactions. Academic chemistry is established, and with its maturity has come an increasingly incurious and risk-averse attitude. So, what's next?

'Business as usual' is not an option. To solve new problems, chemistry must be braver in its research choices and in how it organizes them. As it grew, academic chemistry splintered into many specialized subdisciplines such as organic synthesis, coordination chemistry and laser spectroscopy. This structure worked adequately for the relatively simple problems of the past century, but it will not work for the more complex problems of the next, such as global stewardship of natural resources. The field requires, and is undergoing, a fundamental change.

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

Chemistry rests on three unequal legs: industry, academia and government. Of the three, only universities have the freedom and flexibility to take chemical science in new directions. Industry develops engineered, manufactured products; these should in principle solve societal problems, but must in practice be profitable. Under the short-term constraints of capitalism, industry has largely retreated from long-term research, and mostly focuses on incremental innovation. Government influences research through policy — how funding is allocated between fields, or tax credits for innovation, for instance — but influencing the direction of science is usually incidental to other political agendas. ▶

▶ Universities have fewer constraints.

Current societal problems are too complex for a collection of conservative disciplines. They require a chemistry that is motivated to solve problems about which people care (and for which they will pay). 'Solving practical problems' is often claimed to vulgarize science. It does not: many of chemistry's fundamental discoveries were made in the course of developing practical technologies — catalysis and polymer science, for instance, had their origins in industry. And such problems are often more challenging than the questions fashionable among academic chemists, driven as they are not by unfettered curiosity but by a conservative peer-review system — the spigot that regulates the flow of government funds.

Chemists must remember where the money comes from. Citizens reluctantly allow governments to collect taxes and to spend a minuscule fraction of that money on science, in the faith that by doing so, research will ultimately generate a better world: better health, less conflict, interesting jobs for their children. If that faith falters, the investment will slow or stop.

When funding is flat, the temptation in academia is to retreat to the departmental bunkers. But if chemistry as a discipline is poorly equipped for today's problems, it will wither before the challenges of the future. Reinvention is essential for the continuing relevance and survival of the discipline.

GLOBAL CHALLENGES

Chemistry is the science that connects the relative simplicity of atoms and molecules to the complexity and function of macroscopic matter and of life. Some of the most interesting problems in science, and many of the most important facing society, need chemistry for their solution. Examples include: understanding life as networks of chemical reactions (reactions occurring in water, that most peculiar of solvents); interpreting the molecular basis of disease; global stewardship; the production, storage and conservation of energy and water; and the management of carbon dioxide.

Chemistry — fundamental and applied — has been slow to exploit these research opportunities, and even the best academic chemists have a Prussian-like loyalty to the status quo. Universities that consider themselves the most innovative and radical in their thinking, and that should be in the forefront, are in the rear, defending already familar and well-established fields.

Cracks in the 100-year-old structure of chemistry began to emerge in the 1990s. First, it became clear that chemistry's best intellectual opportunities lay outside its historical boundaries. The new frontiers were the life sciences and materials science. Now others — energy, the environmental sciences, complexity and affordable health care —

offer additional and perhaps even more compelling opportunities.

Second, 'function' replaced 'structure' as the objective. Molecular structure is now easy to define and manipulate in chemistry; function is hard to achieve, especially by design, and usually emerges from empiricism and serendipity. Society does not care if a molecule has a particular structure; it cares if a pill lowers blood pressure, or a panel generates electricity when exposed to sunlight.

Third, academic chemistry is overpopulated. The proliferation of PhD programmes resulted in a demand for research funds that exceeded the (much-expanded) supply, and the imbalance of supply and demand contributed to a peer-review system that protects established fields at the expense of new ideas. These PhD programmes produced too few new ideas and too many average scientists, and neither provided novel solutions to problems (or jobs), nor caught the attention of the public.

Finally, Balkanization of the field has led to specialization in young scientists. If understanding the molecular basis of life requires integrating a high level of expertise in organic reactivity, kinetics, dissipative systems, biochemistry and cell biology, who is to do it?

> *"Society does not care about molecular structure; it cares if a pill lowers blood pressure."*

Probably no single person — and certainly not a young scientist trained narrowly in a technical apprenticeship — has the range of skills demanded.

So far, the field has responded timidly to these pressures. One tentative step is reluctant participation in interdisciplinary programmes and 'centres'. This is a compromise: it recognizes the problem, but does not really solve it. True, centres avoid some of the pitfalls of conventional academic departments, and expose their members to much wider thinking, but reactionary departments remain firmly in control of hiring and promotion. Another is to experiment with new specializations. Departments focused on nanotechnology or biological and medicinal chemistry have sometimes prospered. But exchanging an old specialization for a new one is hardly a sustainable strategy for restructuring a field.

REAL-WORLD SOLUTIONS

Here is what chemists should do instead:

Rewrite the social contract. Chemistry must reorganize to try to solve problems that are important and recognizable to the society that is paying for the research, especially those to do with water, food, health, energy and the environment. To make fundamental discoveries, an approach that starts with practical problems, and uses them to reveal

unsolved fundamental problems, will work at least as well as (and arguably better than) one that starts with the familiar questions of familiar disciplines.

Do away with the old disciplinary structures. Disciplines mature, and must be subsumed into others. Chemistry should cluster its teaching and research around the exciting and uncertain future rather than the ossified historical past. A first step is to merge chemistry and chemical-engineering departments. A second is to form broad new entities that address the most challenging problems that require the skills of chemists. Plausible topics could include functional materials, catalysis, complex dynamic networks, energy, the environment and sustainability, health and out-of-equilibrium systems.

Focus on chemistry's strengths. Chemistry has unique capabilities in many areas: complex kinetics, biological and environmental networks, synthesis of new molecules and forms of matter, examination of the properties of molecules, relating the properties of molecules to the properties of materials, and many others. A focus on these intellectual strengths avoids being second-best in someone else's game.

Teach students, rather than use them. Many subdisciplines of chemistry still use an apprenticeship model in which a professor conceives the problem and strategy, and graduate students execute the bench work. It is hard to imagine a worse way to prepare tomorrow's chemists to work at the integration of many disciplines. Instead, professors should teach students the tools of curiosity. An independent, engaged student, exploring as a colleague in a promising area, will do better work than a simple apprentice.

Chemistry must also change its coursework, to include the hard parts (the role of solvent in chemistry, the importance of thermodynamics in biochemistry, the centrality of mathematics to the study of networks, the subtlety of catalysis and systems of coupled catalysts). It must also include 'non-science' subjects — especially economics and corporate finance and manufacturing — useful in generating practical technologies.

A focus on the practical does not mean ditching fundamental science. It means using fundamental science for a purpose, and practical problems as a stimulus to curiosity. Chemists can still be curious, en route to addressing the big societal challenges of our times. ∎

George M. Whitesides *is the Woodford L. and Ann A. Flowers Institute Professor at Harvard University, Cambridge, Massachusetts 02138, USA.* **John Deutch** *is institute professor of chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, 02139, USA.*
e-mail: gwhitesides@gmwgroup.harvard.edu

What lies ahead

Ten leading chemists set priorities for the forthcoming decades, and reveal the scientists they find inspiring.

PAUL WENDER
Look through a molecular lens

Stanford University, California

Chemistry has often been called a 'central science'. In my view, it is more accurately a 'universal science'. It deals with molecular structure, function and synthesis, subjects of great importance across the whole of science. The problems of our time and of the future are not confined to a single discipline. Indeed, research and how we think about problems are becoming increasingly 'molecularized', because questions require an understanding of atomic-level structure and function and the ability to design and make new molecules and systems — whether drugs, diagnostics, new materials or even functioning cells. From molecular anthropology to molecular zoology (and even molecular gastronomy), we have entered an age of exploration that will lead to transformative innovation.

Those who see problems through a molecular lens are well positioned to address some of the major problems of our time. We cannot hope to improve public health, for example, without a shift in emphasis to early detection and prevention of disease. That in turn requires an understanding of the molecular origins of disease and the design of molecules that can detect early molecular events that lead to disease progression.

Our energy future is also inexorably intertwined with questions of structure and function, whether connected to energy collection, storage or conversion. Smart materials and responsive devices require molecules or molecular systems that both detect an event and structurally change in response to it. We are in the midst of a molecular revolution that will profoundly change our world.

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

CHRISTOPHER C. CUMMINS
Better living through chemistry

Massachusetts Institute of Technology, Cambridge

Chemistry is not always well encapsulated by a simplistic 'big question' formula. Advances in chemistry are needed to address myriad issues, ranging from resource stewardship in the global carbon, nitrogen and phosphorus cycles, to energy-saving ways of doing catalysis with inexpensive, abundant elements. Chemistry has a fundamental part to play in developing a future wherein a high living standard is attainable for all without the sacrifice of our environment and habitat.

I admire Marie Curie (see 'A gallery of greats') tremendously because of the absolute, relentless determination she exhibited in performing the excruciatingly tedious work involved in the isolation of radium. The brilliance of her ideas was backed up by a Herculean work ethic. All budding scientists should read her biography.

MARTYN POLIAKOFF
Make chemical processes greener

University of Nottingham, UK

The key question facing green chemists and process engineers this decade is how to design molecules with specific properties and functions and, then, how to make those molecules with minimal waste and hazard. Chemistry has made huge strides in this in recent years but still we are often reduced to tinkering with molecules and observing the changes in their properties rather like computer hackers probing an unfamiliar piece of software.

Looking further ahead, everything we make and use involves chemical elements and, in the past 100 years, we have squandered many of the planet's resources in making the paraphernalia of modern life. Concentrated deposits of minerals have been mined, and the elements within them scattered across the globe. Now, many of these elements are becoming so scarce that they are 'endangered'. Unless we can invent sustainable substitutes, many of the products that support current society, from laptops to fertilizers, will disappear. And chemists are best placed to make these inventions.

The most important year, rather than person, in chemistry was 1869. Dmitri Mendeleev in St Petersburg proposed the

periodic table and Thomas Andrews in Belfast invented the term 'critical point' at which the distinction between a liquid and a gas disappear. Supercritical fluids have been the focus of my research for the past 25 years and led me into the field of green chemistry. The periodic table has influenced me since high school and, recently, launched me onto YouTube: www.periodicvideos.com.

LAURA KIESSLING
Mimic how nature makes polymers

University of Wisconsin–Madison

One very basic question we must answer is how biological systems make polymers of controlled sequences and lengths without a template. Carbohydrate polymers are the most abundant organic substances on the planet and we do not know how they are generated and how their lengths are controlled. This information could allow us to better harvest cellulose for energy, design better vaccines for pathogens, control growth-factor signalling pathways in cancer or development, and devise new types of antimicrobials. Understanding how nature makes polysaccharides could also provide insight into a wide range of polymerization reactions, including those that underlie the formation of telomeres —protective caps on chromosomes — and their role in cancer.

I admire so many chemists, living and dead. Pushed to choose one, I would go with Emil Fischer (1852–1919) for his imaginative applications of organic synthesis to address problems in biology. He was perhaps the first chemical biologist.

E. W. 'BERT' MEIJER
Foster synthetic self–assembly

Eindhoven University of Technology, the Netherlands

'How far can we push chemical self-assembly?' is without doubt the most intriguing challenge we as scientists have to solve in the decades to come. Synthetic chemists are able to prepare almost every molecule present on Earth, but we now need sophisticated ways to assemble and organize these molecules into functional molecular objects. New methodologies, mechanistic insights, out-of-equilibrium and kinetically controlled assemblies are just a few important topics in this interdisciplinary field of science.

I admire a perfect blend between Jacobus van 't Hoff (1852–1911) and Hans Wynberg. Van 't Hoff, the first chemistry Nobel laureate in 1901, proposed seminal insights into the three-dimensional arrangements of atoms in space, thereby introducing the field of stereochemistry. Remarkably, he got this idea at the age of 22, showing the strength of combining a passion for science with originality of thought. I admire my PhD adviser Hans Wynberg for introducing me to the fascinating world of chiral molecules and for his drive to stimulate the young generation to work at the frontiers of science.

PAUL ALIVISATOS
Replicate photosynthesis

Director, Lawrence Berkeley National Laboratory, California

This will be the decade when we finally learn how to make artificial photosynthesis work in a practical way. The goal dates back to Melvin Calvin (1911–97), who developed our understanding of the biological carbon cycle, and who appreciated the need to establish a stable cycle for human use of energy. An artificial photosynthetic system could provide us with a sustainable form of energy for the future. This grand challenge requires chemists to solve many deep and long-standing problems. For instance, we need to understand multi-electron and multi-step catalytic events more deeply, so that we can design better catalysts for oxygen generation from water and for the reduction of carbon dioxide to fuel. We need to learn how to assemble precise multi-component nanoscale light-absorbing and charge-separation systems and to integrate these with catalysts. These systems need to be grown from abundant materials, by processes that can be scaled to vast areas, by inexpensive means.

This problem was investigated in the late 1970s and early 1980s, but then there was a 30-year hiatus. In the intervening decades, nanoscience has developed, and there are new theoretical and analytical tools at our disposal.

Rather than pick a favourite scientist, I prefer to call attention to a remarkable paper: 'On the influence of carbonic acid in the air upon the temperature of the ground' (*Philos. Mag.* **41,** 237–276; 1896). In it, the physical chemist Svante Arrhenius (1859–1927) estimates that a doubling of the carbon dioxide in the atmosphere will increase the earth's temperature by 5 °C. A few years later he refined his calculation to get 2.1 °C. At the time, he considered it

unlikely that human activity could lead to such an increase. Now more than 100 years have passed, the scale of human activity has grown enormously, and we can see that Arrhenius got it mostly just right.

KAREN WOOLEY
Enhance selective interactions

Texas A&M University, College Station

Over the next decade the design and study of polymers as functional materials for medical and other applications must address three primary challenges. First, how to enhance selective interactions while avoiding non-selective ones, so that molecules can target specific tissues *in vivo*. Second, how can chemists create single, well-defined structures, as formed in nature, instead of populations of materials with varying composition, structure and size that result from experiments in the lab? Third, how can synthetic organic chemists extend the exquisite control they wield over the construction of natural drug products and analogues to allow new synthetic targets for chemical manipulation, including ribosomes and viruses. All of these require greater control over intermolecular interactions.

As a young girl, I admired Marie Curie, but over the years, my appreciation for several 'giants' of modern chemistry has grown, especially for those, such as Robert Grubbs (who shared the 2005 Nobel Prize in Chemistry for his work on organic reactions), who are driven by scientific curiosity and maintain a humble, friendly personality and active mentorship of young chemists.

DAVID KING
Solar power is the future

Director, Smith School of Enterprise and the Environment, Oxford, UK

The next decade will hopefully see the generation of an efficient photovoltaic material that can be cheaply produced and is attractive to architects and builders to use on the outside of buildings — in plastics, paint or ceramics, for example — thereby revolutionizing the use of solar power.

The figure I find inspiring is Antoine Lavoisier (1743–94), the father of modern chemistry. The major intellectual contribution contained in his outstanding monograph *Elementary Treatise on Chemistry* (1789) marked the development of an
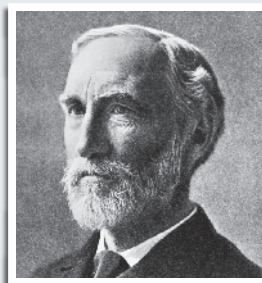
A GALLERY OF GREATS

The work of these scientists, among others, has inspired the current generation of chemists. The International Year of Chemistry 2011 also celebrates the centenary of Marie Curie's Nobel Prize in Chemistry.
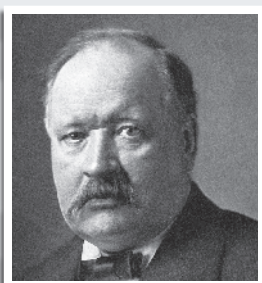
Antoine Lavoisier
(1743–94)
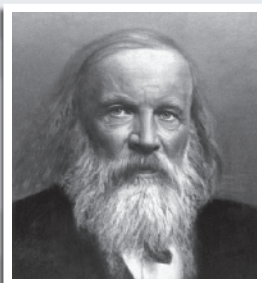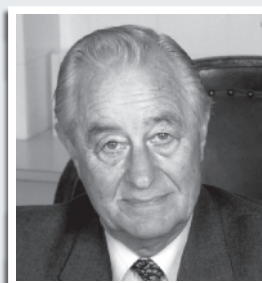
Dmitri Mendeleev
(1834–1907)

Josiah Willard Gibbs
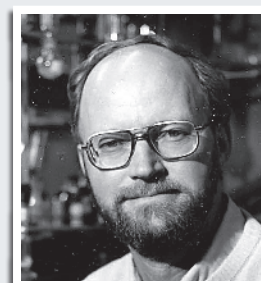(1839–1903)

Jacobus van 't Hoff
(1852–1911)

Svante Arrhenius
(1859–1927)

Marie Curie
(1867–1934)

George Porter
(1920–2002)

Robert Grubbs
(1942–)

understanding of chemical elements, the law of conservation of mass, the end of the phlogiston theory (which had been blocking progress) and even the end of alchemy. Modern chemistry, materials science and physics emerged, to a large extent, from the progress that this work represented.

JOANNA AIZENBERG
Promote sustainable living

Harvard University, Cambridge, Massachusetts

Chemists are tackling questions of sustainability and will continue to do so. They are following several important and interrelated avenues: development of efficient alternative energy sources, smart and sustainable use of natural resources, preserving the environment, combating hunger and improving human health and living standards across the globe. The processes and materials needed in all these areas, such as effective use of solar energy, hydrogen-based fuel, nuclear-waste minimization and handling, water purification, adapting coal-based technologies to the highest environmental standards and

developing more-productive crops, will all require the contribution of chemists — organic and inorganic, physical, polymer, materials and biochemists.

I most admire Josiah Willard Gibbs (1839–1903). He contributed immensely to establishing chemical thermodynamics and physical chemistry. His work had enormous breadth for his time, for any time, spanning mathematics, physics, chemistry and engineering. The list of phenomena and laws he discovered, studied and explained, and that carry his name, is so impressive that even a fraction of them would make their author a very prominent scientist, worthy of remembering for many years. Where would chemistry be today without all that Gibbs energy?

GRAHAM FLEMING
Catalyse energy production

Vice-chancellor for research, University of California, Berkeley

The major problems for chemists to help solve ('important' problems can no longer be solved by any single discipline) all relate to energy, its generation and its storage.

Chemists can contribute key advances in the creation of batteries with twice the energy density and five times lower cost (so ten times better) than current batteries. This is required to make electric cars a reality. Chemists need to develop better catalysts for conversion of solar energy to liquid fuels and to exploit new chemical cycles (that don't currently exist in nature) for creating useful energy, using materials abundant on Earth. They need to work with biologists to radically improve the efficiency of biological photosynthesis by reprogramming the natural machinery for both food and fuel production.

Long term, chemists need to develop ways of synthesizing molecular systems that can regulate their own behaviour in response to external changes, and that can repair themselves if damaged. This latter task is probably too difficult to achieve in ten years, but should still be a goal.

The chemist I admire is my thesis adviser George Porter (1920–2002) for his ability to do groundbreaking science (leading to a Nobel prize), his ability to connect with young people and non-scientist adults, and his strong sense of social justice. As I prepared his biographical memoir for the Royal Society, I learned how much effort George had put in to securing the release and exit visas for 'refuseniks' in the Soviet Union. ■
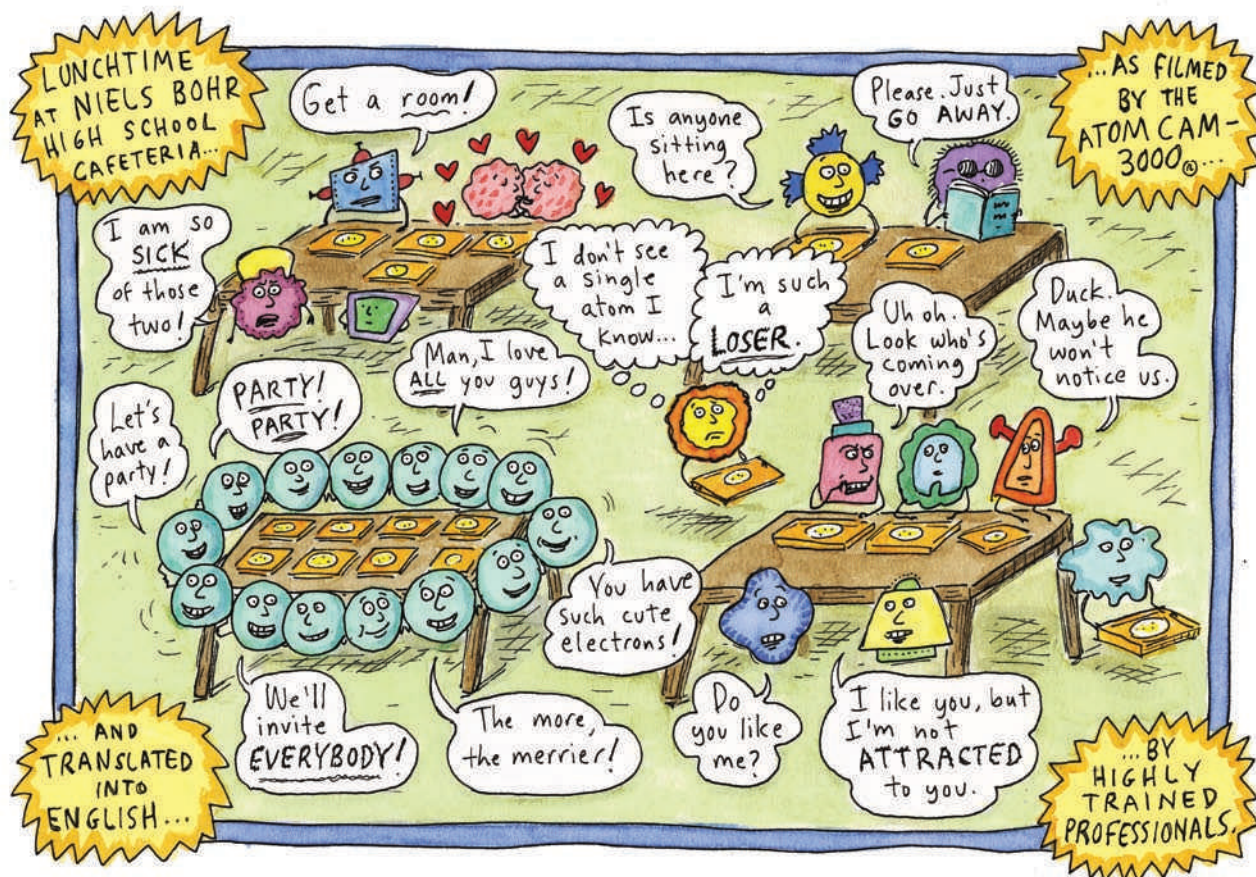
Beyond the bond

More than ever before, new techniques show the bond to be a convenient fiction, albeit one that holds the field of chemistry together, finds **Philip Ball.**

Not so long ago, the chemistry student's standard text on the theory of chemical bonding was Charles Coulson's *Valence* (1952). Absent from it was this theoretical chemist's real view on the sticks that generations of students have drawn to link atoms into molecules. "A chemical bond is not a real thing: it does not exist: no one has ever seen it, no one ever can. It is a figment of our own imagination," he later wrote[1].

There is a good reason for postponing this awkward truth. The bond is literally the glue that makes the entire discipline cohere; to consider it an objective reality is necessary for any kind of discourse on chemistry. The discipline is in fact riddled with such convenient (and contested) fictions, such as electronegativity, oxidation state, tautomerism and acidity.

Disputes about the correct description of bonding have ruffled chemists' feathers since the concept of molecular structure first emerged in the mid-nineteenth century. Now they are proliferating, as new theoretical and experimental techniques present fresh ways to probe and quantify chemical bonds[2]. Traditional measures such as crystallographic interatomic distances and dissociation energies have been supplemented by spectroscopic techniques for determining vibrational frequencies, methods such as nuclear magnetic resonance to measure shifts in the electronic environment of atoms and their magnetic interactions, measurements of force constants (bond stiffness) and a host of quantum-chemical tools for calculating such aspects as electron distributions or localization.

An indication of the shifting ground is the recent decision to redefine the nature of the hydrogen bond in the light of experimental results that alter the traditional electrostatic description.

The nature of the chemical bond is now further complicated by the introduction of

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

the dynamical dimension. Molecules have traditionally been regarded, if not as static, then as having clear architectural frameworks that are merely shaken and rotated by thermal motions. The bonds get stretched and bent, but they still have an equilibrium length and strength that seems to justify their being pictured as lines and stalks. Now, thanks to ultrafast spectroscopies, such time-average values may not always accurately characterize either structure or reactivity. What is measured in a bond depends not just on how but on when it is measured.

Some chemists argue that as a result of this, the very existence (or not) of a bond depends on how the problem is probed. Others are committed to absolute criteria. This difference of opinion goes to the heart of what chemistry is about: can all be reduced to quantum physics, or are fuzzy rules of thumb essential? More pressingly, the issue of how best to describe a chemical bonding pattern has tangible implications for a wide range of problems in chemistry, from molecules in which atoms are coerced out of their

usual bonding geometry to the symmetric hydrogen bond (in which a bound hydrogen atom is shared equally between two other atoms), and new variations on old themes such as aromaticity (special patterns of 'delocalized' bonds, like those in benzene)[3].

Just about every area of chemistry harbours its own bonding conundrums. Most illustrate that we have a far from exhaustive understanding of the ways in which quantum rules will permit atoms to unite — and that in consequence, the synthetic inventiveness of chemists suffers from a limited view of the possibilities.

CARVING UP ELECTRONS

We can all agree on one thing: chemical bonding has something to do with electrons. Two atoms stick together because of the arrangement of electrons around their nuclei. In the nineteenth century, it was commonly thought that this attraction was electrostatic: that atoms in molecules are positively or negatively ionized. That left the puzzle of how identical atoms can form diatomic molecules such as H_2 and O_2. American chemist G. N. Lewis proposed that bonding can instead result from the sharing of electrons to create filled shells of eight, visualized as the corners of a cube.

In the 1920s and 1930s another American chemist, Linus Pauling, showed how this interaction could be formulated in the language of quantum mechanics as the overlap of electron orbitals (see *Nature* **468,** 1036; 2010). In essence, if two atomic orbitals each containing a single electron can overlap, a bond is formed. Pauling generalized earlier work on the quantum description of hydrogen to write an approximate equation for the wavefunction created by orbital overlap. This became known as the valence-bond (VB) description.

But an approximation is all it is. Around the same time, Robert Mulliken and Friedrich Hund proposed another approximate wavefunction, which led to an alternative way to formulate bonds: not as overlaps between specific orbitals on separate atoms, but as electron orbitals that extend over many atoms, called molecular orbitals (MOs). The relative merits of the VB and MO descriptions were debated furiously for several decades, with no love lost between the protagonists: Mulliken's much-repeated maxim, "I believe the chemical bond is not so simple as some people seem to think", was possibly a jibe at Pauling. By the 1960s, for all Pauling's salesmanship, MO theory was generally agreed to be more convenient for most purposes. But the debate is not over[4], and Roald Hoffmann of Cornell University in Ithaca, New York, insists that "discarding any one of the two theories undermines the intellectual heritage of chemistry".

Both options are imperfect, because they insist on writing the electronic wavefunction

as some combination of the wavefunctions of individual electrons. That's also the basis of the Hartree–Fock method for calculating the wavefunction and energy of the lowest-energy state (ground state) of a molecular system — a method that became practical in the 1950s, when computers made it possible to solve the equations numerically. But separating the wavefunction into one-electron components is a fiction, as they all influence one another: the behaviour of one electron depends on what all the others are doing. The difference between true ground-state energy and that calculated using the Hartree–Fock approach is called the correlation energy. More recent computational methods can capture most of the correlation energy, but none can give an exact solution. As a result, describing the quantum chemical bond remains a matter of taste: all descriptions are, in effect, approximate ways of carving up the electron distribution.

If that were the limit of the bond's ambiguity, there would be little to argue about. It is not. There is, for example, the matter of when to regard two atoms as being bonded at all. Pauling's somewhat tautological definition rather gave the game away: a group of atoms is to be considered bonded when it is "convenient for the chemist to consider it as an independent molecular species". Pauling admitted some of the consequent ambiguities: for example, although his
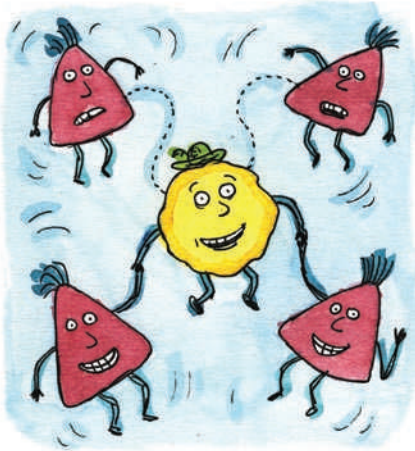
definition in general excludes the weak van der Waals attraction that occurs between all atoms, occasionally — as in the association of two oxygen molecules into the O_4 cluster — even this force can be strong enough to be regarded as a chemical bond.

It's no use, either, to suggest (as Coulson did) that a bond exists whenever the combined energy of the atoms is lower than that when they are separated by an infinite distance. For this is essentially always the case**,** at least for electrically neutral atoms. Even two helium atoms experience mutual van

der Waals attraction, which is why helium is a liquid at very low temperatures; but they are not generally thought to be chemically bonded as a result.

Besides, the 'bonded or not' question becomes context-dependent once atoms are embedded in a molecule, where they may be forced close together merely by the atoms around them, and where there is inevitably some arbitrariness in deciding which electrons 'belong' to which atoms. The resulting ambiguities were illustrated recently when three experts failed to agree about whether two sulphur atoms in an organometallic compound are linked by a bond[5]. The argument involved different interpretations of quantum-chemistry calculations, tussles over the best criteria for identifying a bond, and evidence of precedent from comparable compounds.

"Chemists' inventiveness suffers from a limited view of the possibilities."

All this is merely a reminder that a molecule is ultimately a set of nuclei embedded in a continuous electron cloud that stabilizes a particular configuration, which balls and sticks can sometimes idealize and sometimes not. But disputes about the nature of the chemical bond are not simply semantic. It matters, for example, whether we regard a very strong multiple bond as quintuple or sextuple, even if this is a categorization that only textbooks, and not nature, recognize — not least, because textbook concepts are what provide a discipline with intellectual coherence and consistency.

Besides, how we talk about bonds can determine our ability to rationalize real chemical behaviour. For example, the different descriptions of the bonds in what are now called non-classical ions of hydrocarbons — whose relative merits were furiously debated in the 1950s and 1960s — have direct implications for the way these species are predicted to react. Whether to consider the bonding non-classical, involving electrons spread over more than two atomic nuclei, or tautomeric, involving rapid fluctuations between conventional two-atom bonds, had immediate consequences for organic chemistry.

Perhaps one might seek a distinction between bonded and not-bonded in terms of how the force between two atoms varies with their separation? Yes, there is an exponential fall-off for a covalent bond such as that in H_2, and a power-law decay for van der Waals attraction. But the lack of any clear distinction between these two extremes has been emphasized in the past two decades by the phenomenon of aurophilicity[6], in which gold atoms that have only a few chemical groups attached to them in organometallic compounds tend to aggregate, forming dimers
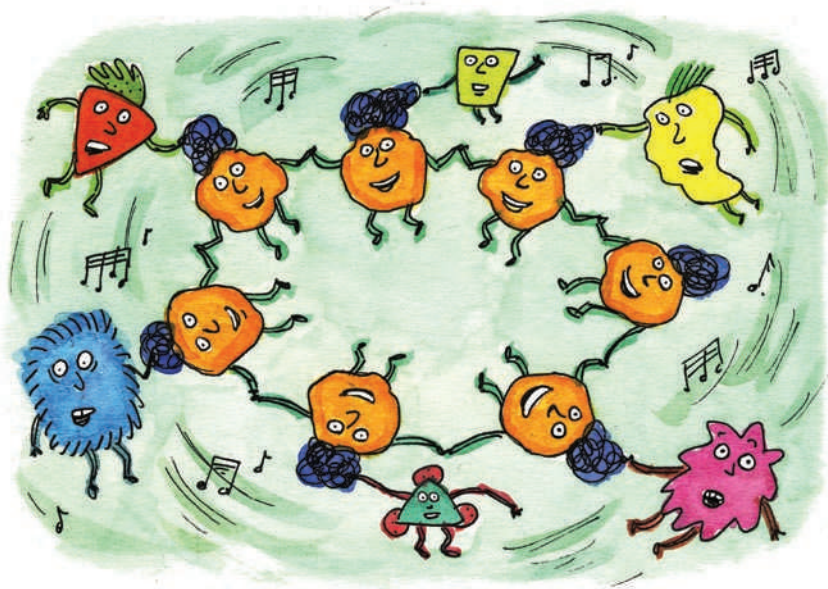
and linear chains. The basic interaction in these aurophilic bonds between gold has the same origin as the van der Waals force: the electron clouds 'feel' each other's movements, so that random fluctuations of one induce mirror-image fluctuations of the other. But that interaction is modified here by relativistic effects: changes in electron energies resulting from their high speeds around gold's highly charged, massive nuclei[6,7]. Aurophilic bonds have therefore been described as a 'super van der Waals' interaction. Does that make them true bonds? It's chemically meaningful to treat them that way (they'll even serve for cementing new 'designer' molecular crystals[8]), but perhaps at the cost of relinquishing another potential criterion for 'real' bonds.

MOVING TARGET

Perhaps the biggest challenge to the simple notion of a chemical bond is dynamics. Atomic motions make even a 'simple' molecule complex; any movement of one nucleus demands that the entire electron cloud adjusts. A jiggle of one group of nuclei can make it easier to cleave off another.

This complication never used to matter much in chemistry, because the movements were too rapid to be observable, much less exploitable. Ultrashort pulsed lasers moved the goal posts. For example, energy can be pumped into a vibrational mode to weaken a specific bond, enabling selective molecular surgery[9]. Chemists can ask about the chemical behaviour of a molecule at a particular moment in its dynamical evolution. Even a strong bond is weakened when a vibration stretches it beyond its average, equilibrium length, so in ultrafast chemistry it may no longer be meaningful to characterize bonds simply as strong or weak. As Fleming Crim of the University of Wisconsin–Madison puts it: "A bond is an entity described by quantum mechanics but not a fixed 'entity'

in that it will behave differently depending on how we perturb and interrogate it." The trajectory of a chemical reaction must then be considered not as a making and breaking of bonds but as an evolution of atoms on a potential-energy surface. This was always implicit in classical drawings of transition states as molecular groupings containing dashed lines, a kind of 'almost bond' in the process of breaking or forming. Now that picture is explicitly revealed as a mere caricature of a complicated dynamical process in space and time.

Underlying most of these discussions is an unspoken assumption that it is meaningful to speak, if not of a 'bond' as an unchanging entity, then at least of an instantaneous bound state for a particular configuration of nuclei. This assumes that the electrons can adjust more or less instantly to any change in the nuclear positions: the Born–Oppenheimer approximation. Because electrons are so much lighter than nucleons, this assumption is usually justified. But some clear breakdowns of the approximation are now well documented[10]. They are best known in solid-state systems, and in fact superconductivity is one of the consequences, resulting from a coupling of electron and nuclear motions. Such things may also happen in molecules, particularly in the photochemistry of poly-atomic molecules, which have a large number of electronic states close in energy[11]; they have also been observed for simple diatomic molecules in strong electric fields[12]. As a result, the molecular degrees of freedom may become interdependent in strange ways: rotation of the molecule, for example, can excite vibration. In such situations, the very notion of an electronic state begins to crumble[10].

> *"The biggest challenge to the simple notion of a bond is dynamics."*

These advances in dynamical control of quantum states amount to nothing less than a new vision of chemistry. The static picture of molecules with specific shapes and bond strengths is replaced by one of a bag of atoms in motion, which can be moulded and coaxed into behaviours quite different from those of the equilibrium species. Such a vision does not demand that we abandon old ideas about chemical bonds, nor does it truly challenge the ability of quantum theory to describe atoms and their unions. It recommends that we view these bonds as degrees of attraction that wax and wane — or as cartoon representations of a molecule's perpetual tour of its free-energy landscape.

At a meeting in 1970, Coulson asserted that the simple notion of a chemical bond had already become lost, and that "something bigger" was needed to replace it — a subject, he suggested, for another meeting 50 years hence. That moment is almost upon us.

Yet we needn't fret that the 'rules' of bonding are up for grabs — quite the reverse. Although there may be some parts of science fortunate enough to be exhaustively explained by a single, comprehensive theory, this isn't likely to be a general attribute. We are typically faced with several theories, some overlapping, some conflicting, some just different expressions of the same thing. Our choice of theoretical framework might be determined less by the traditional criterion of consistency with experiment than by subjective reasons. According to Hoffmann, these preferences often have an aesthetic component, depending on factors such as simplicity, utility for telling a story about chemical behaviour, the social needs of the community and whether a description is productive.

As Hoffmann says: "Any rigorous definition of a chemical bond is bound to be impoverishing." His advice to "have fun with the fuzzy richness of the idea" is well worth heeding. ∎

Philip Ball *is a writer based in London.*

1. Coulson, C. A. *The Spirit of Applied Mathematics* 20–21 (Clarendon Press, 1953).
2. *J. Comput. Chem.* special issue **28,** 1–466 (2007).
3. Cortés-Guzmán, F. & Bader, R. F. W. *Coord. Chem. Rev.* **249,** 633 (2005).
4. Hoffmann, R., Shaik, S. & Hiberty, P. C. *Acc. Chem. Res.* **36,** 750–756 (2003).
5. Alvarez, S., Hoffmann, R. & Mealli, C. *Chem. Eur. J.* **15,** 8358–8373 (2009).
6. Pyykkö, P. *Chem. Soc. Rev.* **37,** 1967–1997 (2008).
7. Schmidbaur, H., Cronje, S., Djordjevic, B. & Schuster, O. *Chem. Phys.* **311,** 151–161 (2005).
8. Katz, M. J., Sakai, K. & Leznoff, D. B. *Chem. Soc. Rev.* **37,** 1884–1895 (2008).
9. Crim, F. F. *Science* **249,** 1387 (1990).
10. Sukumar, N. *Found. Chem.* **11,** 7–20 (2009).
11. Worth, G. A. & Cederbaum, L. S. *Ann. Rev. Phys. Chem.* **55,** 127–158 (2004).
12. Sindelka, M., Moiseyev, N. & Cederbaum, L. S. preprint at www.arxiv.org/abs/1008.0741 (2010).

Lauren Redniss's cyanotype drawings of the Curies symbolize radium's faint blue glow.

HISTORY

Radioactive romance

Giovanni Frazzetto is captivated by an illustrated biography of Marie and Pierre Curie.

Writer and artist Lauren Redniss's *Radioactive* is no ordinary biography of Marie and Pierre Curie. The story of radioactivity, one of the most exciting discoveries of the past 100 or so years, is brightly visualized through Redniss's imagination in her illustrated book. Ideas, scientific choices, motivations and insatiable passions unfurl in her elegant cyanotype drawings and are enacted by ethereal figures set into motion by the author's eloquence.

The tale begins with a taste of Pierre and Marie's lives before their magnetic meeting, charting Marie's arrival in Paris from Poland in 1891 and her integration into the French academic structure. It proceeds with their early studies on radiation, their joint Nobel Prize in Physics (with Henri Becquerel) in 1903, and Pierre's death in a street accident in 1906. The pages tell of Marie's affair with her late husband's brilliant former student, Paul Langevin, and of the ensuing scandal, which broke out at the time Marie was awarded her second Nobel prize in 1911 for the isolation of polonium and radium (the 100th anniversary of which we celebrate this year).

Redniss unfolds plot and content with skill and kindness. She introduces suspense and selects affectionate, unforgettable details. The concepts of radioactivity, half-life and nuclear fission are delivered with poetic ease, and used as metaphors to parallel steps in the protagonists' lives. The result is a

Radioactive: Marie and Pierre Curie, A Tale of Love and Fallout
LAUREN REDNISS
It Books: 2010.
208 pp. $29.99

Radioactive
Artwork from the book is on display at the New York Public Library, 14 January–17 April 2011.

linguistically rich and visually captivating portrait of two formidable scientific figures and their extraordinary discoveries.

Radium is one of nature's most bewitching elements. The radiation that stems from it evoked wonder in the discoverers themselves, who gave lyrical descriptions of the spontaneous and luminous energy of the substances they had isolated: "These gleamings, which seemed suspended in the darkness, stirred us with ever new emotion and enchantment … the glowing tubes looked like faint, fairy lights."

The thread of discovery is intertwined with contemporary applications throughout the book. We learn how Pierre's early studies of piezoelectricity are applied today in measuring caffeine levels in beverages, in treating hypothermia and in propelling droplets in inkjet printers. The implications of the Curies' science also had a dark side. We are reminded of the discoveries leading up to the atomic bomb, of the Chernobyl disaster and how former KGB officer Alexander Litvinenko was apparently killed by polonium poisoning in London in 2006.

Cyanotype, the printing process used by Redniss for many of the images in the book, is a deliberate choice. A negative transparency of each illustration is placed on paper coated with chemicals that, on exposure to sunlight, reduce to become the dye Prussian blue. The dark areas in the negative remain clear. For Redniss, the derivation of the final drawing from its negative symbolizes radium's faint turquoise glow. The technique also captures Marie's simultaneous fragility and strength: cyanotype prints fade when exposed to light, but can be rejuvenated by storing them in the dark. And Prussian blue capsules are an approved treatment for radioactive contamination in humans.

Radioactive is romance, history, art and poetry woven together into a unique style of biography. Redniss introduces us to the intense and sometimes frail lives of brilliant devoted scientists, displaying her enchantment with the lyrical aspects of science. ∎

Giovanni Frazzetto *is at the BIOS Centre of the London School of Economics, UK, and a fellow at the ICI Berlin Institute for Cultural Inquiry, Germany.*
e-mail: g.frazzetto@lse.ac.uk

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

Far from defining the chemical elements, Robert Boyle in fact doubted their existence.

IN RETROSPECT
The Sceptical Chymist

Robert Boyle's widely misunderstood book elevated the status of chemistry, explains **Lawrence Principe**.

It can be said of books, as Shakespeare said of people, that "some are born great, some achieve greatness, and some have greatness thrust upon them". Robert Boyle's *The Sceptical Chymist* (1661) falls into the last category. Widely celebrated as a landmark, it remains misrepresented, misunderstood and unread. It has been praised for opening the way to modern chemistry by sweeping away misguided alchemy and for providing the first modern definition of an element. Yet it did neither. Both Boyle and the book are important, but not for the reasons often cited.

Boyle (1627–91), seventh son of the richest man in Britain, devoted his life and fortune to the study and advancement of science and Christianity. At his sister's Pall Mall address in London where he lived much of his adult life, he performed countless experiments, received visitors and colleagues, and wrote many books. *The Sceptical Chymist* was Boyle's fourth volume, following a devotional treatise, a collection of chemical essays and an

The Sceptical Chymist
ROBERT BOYLE
First published 1661.

account of experiments performed with the 'Pneumatick Engine', or air-pump. That device, built for him by Robert Hooke, led to the enunciation of the inverse relationship between the pressure and volume of gases, later named Boyle's law.

The Sceptical Chymist is not an easy read. Begun as a dialogue among proponents of various chemical systems, the book is an unpolished pastiche of several unfinished works. It is prolix, repetitive, disjointed and occasionally contradictory.

Boyle's goals were to question some commonly held chemical theories, to make chemical practitioners more "philosophical" and less commercially focused, and to

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

raise chemistry's status. In his day, the field was held in low esteem; it had no place at universities, its practice was dirty, smelly and laborious. Chemical techniques were used mainly to produce medicaments, a situation preserved in the British usage of the word chemist to mean pharmacist. Chemical textbooks were predominantly lists of pharmaceutical preparations. Boyle believed that chemistry should be much more. For him, it was the key to understanding nature.

Boyle held that chemical pharmacists and textbook writers — the vulgar chymists as he called them — relied on unsound ideas. Chief among them was the concept, advanced by the sixteenth-century medical reformer Paracelsus, that all substances were composed of three essential components. Paracelsians based this belief on their finding that organic materials could be distilled to provide volatile, inflammable and saline fractions, which they named Mercury, Sulphur and Salt, respectively. Boyle argued that these could not be isolated from all materials, and were produced rather than separated by the distilling fire. Hence, they could not be universal principles. Moreover, when the vulgar chymists thought they were concentrating the 'essence' of something by distillation, they were reducing its therapeutic value through thermal decomposition.

Boyle drew on a long chemical tradition for his arguments. A generation earlier, Joan Baptista van Helmont, one of the most influential chemists of the seventeenth century, had expressed the same sceptical views about fire analysis and the three principles. Indeed, contemporaneous readers viewed *The Sceptical Chymist* as a popularization of van Helmont, who proposed that all substances were produced by modification of a single primordial principle — water. Boyle likewise attributed everything to a single stuff, which he called Catholick matter. Differences in the shape, size and motions of minute particles of this matter gave rise to the diverse substances of the world, and changing these characteristics using chemistry could transform any material into any other one. So rather than defining chemical elements as we know them, Boyle doubted that there were any. His view, moreover, supported the alchemists' goal of transmuting base metals into gold.

Although misrepresented by later generations as magical or fraudulent, alchemy was in Boyle's day a serious means of investigating nature. Alchemy had an important and positive role in the history of science. Boyle first learned chemical ideas and practices (and studied van Helmont) under the tutelage of George Starkey, an immigrant from America and graduate of Harvard College. Using the pen-name Eireneaus Philalethes, Starkey wrote influential books on transmutation and devised a sophisticated theory of matter. Boyle himself strove to prepare the philosophers' stone, the secret substance

able to transmute metals; he recounted how he had witnessed a demonstration of its powers and tested the gold it produced. He even petitioned Parliament successfully in 1689 to repeal a law forbidding gold-making, which he thought impeded research towards discovering the secret of the stone.

Far from repudiating alchemy, *The Sceptical Chymist* cites alchemical texts and theories to criticize the vulgar chymists. In the 1680 second edition, Boyle referred respectfully to "a much higher order" of chemists "able to transmute baser Metalls" and do things that less-skilled chemists "have judg'd impossible". Although Boyle expressed frustration at the secrecy of alchemical authors, he also thought of them as preferable to and possessed of greater knowledge and expertise than the "meer operators" against whose "unphilosophical" practice his book was directed.

The Sceptical Chymist was neither the most widely read nor the most important of Boyle's works. Its fame was thrust upon it by retrospective attempts more than two centuries after its publication to locate a revolutionary moment for chemistry. This view attributed to Boyle innovations that later generations thought were important — such as the definition of an element and the repudiation of alchemy — but which were far from Boyle's mind. Those attempts worked only because few cared to wade through Boyle's prose and still fewer took time to recognize his motives. Besides, such revolutionary moments exist mostly in the minds of those who imagine that science develops by sudden leaps made by isolated geniuses rather than by slow, laborious steps accumulated and shared by many talented workers.

> **"Boyle's view supported the alchemists' goal of transmuting base metals into gold."**

The Sceptical Chymist was one facet of Boyle's larger project to elevate the status of chemistry, to free it from servitude to medical and commercial endeavours, and to use it for exploring and explaining the hidden workings of nature. Today, chemistry continues to serve the basic physical needs of society and the intellectual and methodological needs of other sciences more than any other discipline. Even as it continues to struggle with its identity in the ever-changing landscape of modern science, chemistry has become established as an independent, fundamental and philosophical discipline. Boyle, although sceptical of elements and convinced of alchemy, played a key part in this achievement. ∎

Lawrence Principe *is Drew Professor of the Humanities at Johns Hopkins University, Maryland, USA. He is author of* The Aspiring Adept, Robert Boyle and His Alchemical Quest.
e-mail: Lmafp@jhu.edu

Books in brief
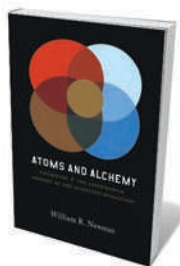
Atoms and Alchemy: Chymistry and the Experimental Origins of the Scientific Revolution
*William R. Newman* UNIVERSITY OF CHICAGO PRESS *235 pp. $36 (2011)*
The alchemical roots of chemistry are laid bare by historian of science William Newman. Countering the view that these medieval practices were unscientific, he argues that alchemists seeded ideas such as the particulate nature of matter, the mechanical view of the Universe and the concept of the elements. By discussing how seventeenth-century chemist Robert Boyle was influenced by his predecessors, Newman demonstrates how alchemy helped rather than hindered the emergence of modern science.
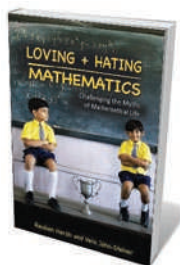
Loving and Hating Mathematics: Challenging the Myths of Mathematical Life
Reuben Hersh and Vera John-Steiner PRINCETON UNIVERSITY PRESS *428 pp. $29.95 (2011)*
Mathematics gets a bad press. Its practitioners are often portrayed as aloof hyper-rationalists with oodles of logic but no emotional intelligence. Not so, say mathematician Reuben Hersh and linguist Vera John-Steiner. Focusing on the emotional side of the discipline, they reveal mathematicians' passions, collaborations and love affairs. The stories range from those who sought solace in equations to some who were driven to murder by a maths obsession.
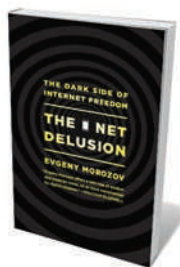
The Net Delusion: The Dark Side of Internet Freedom
*Evgeny Morozov* PUBLICAFFAIRS *432 pp. $27.95 (2011)*
The Internet is often said to be synonymous with democracy and freedom. But there is a dark side, cautions journalist Evgeny Morozov. Rather than liberating citizens worldwide from oppressive rule, digital technology can be just as powerful in suppressing free speech, enabling surveillance and in pacifying populations through immersive entertainment. He calls for an end to glib assumptions that the Internet is inherently good, and for increased efforts towards digital diplomacy, which he says needs as much oversight and consideration as any other kind.
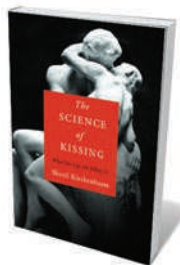
The Science of Kissing: What Our Lips Are Telling Us
Sheril Kirshenbaum GRAND CENTRAL PUBLISHING *272 pp. $19.99 (2011)*
We really do remember our first kiss, explains science writer Sheril Kirshenbaum in her palatable book about why we pucker up. Weaving together evolutionary biology, psychology and anthropology, she describes how different cultures do it differently, and reveals how we sense the fitness of our kissing partner while enjoying the neurological hit of close sensory contact. She explains how men and women kiss differently and why kissing disappeared during the Dark Ages.
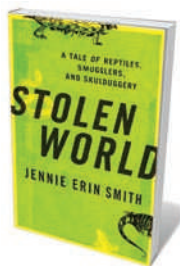
Stolen World: A Tale of Reptiles, Smugglers, and Skulduggery
Jennie Erin Smith CROWN *336 pp. $25 (2011)*
The alien appeal of alligators, snakes and lizards leads enthusiasts to scour the world for rare reptile species. Some collectors will go to great lengths to get them, resulting in a multimillion-dollar black market in illegally imported animals. Through interviews with knife-wielding reptile dealers, science reporter Jennie Smith uncovers this bizarre underworld. She finds that not even zookeepers are exempt from pushing the limits of morality to obtain an unusual breed.

Fossil finds such as this whale — assembled by explorer Jacques Cousteau in Antarctica in 1979 — have helped to trace the complex ancestry of vertebrates.

PALAEONTOLOGY

In the bones

Jan Zalasiewicz enjoys a romp through vertebrate evolution and its eccentric scholars.

Brian Switek's *Written in Stone* boasts a fine selection of the eccentric characters that grace the annals of palaeontology. Some are well known, such as William Buckland and his gluttonous appetite for both collecting and eating animals. Less familiar figures include Baron Franz Nopcsa von Felső-Szilvás — a Hungarian aristocrat, spy and interpreter of pterosaur flight — and the impresario Albert Koch, who assembled sea monsters for the goggle-eyed public in the mid-nineteenth century. These characters add spice to a narrative that takes us, once again, in search of our origins.

Switek's book joins a seemingly inexhaustible supply of evolutionary tomes. Most have been generated in response to the fact that some people refuse to accept that the human species is linked by descent with the rest of the living world — and has shared ancestors with the gorilla and the chimpanzee in recent geological time. One more book will not change those beliefs. But Switek's engaging account may tempt the uncommitted to appreciate how interesting is the underground

Written in Stone: Evolution, the Fossil Record, and Our Place in Nature
BRIAN SWITEK
Bellevue Literary Press: 2010. 320 pp. $17.95

world, and how the vast storehouses of Earth's strata further our understanding of how life developed.

This is not a book about the small fry of Earth, despite the spectacular ammonites on the cover. That is a pity, because ammonites, graptolites and foraminifera are abundant visceral reminders of the reality of organic change through time. Rather, Switek has been fascinated by the bones of vertebrates, large and small, since his youthful pilgrimages to the dinosaur exhibits of the American Museum of Natural History in New York. It is they who stalk the pages. Their remains are rare and fragmentary, but humans empathize more with dinosaurs and dodos than with snails and scallops.

The book begins with the cautionary tale of Ida, the beautiful million-dollar ancient lemur that became a lesson in how not to claim human ancestry for a fossil. Ida also brings the reader into gentle contact with cladistics and phylogenetic relationships. Next comes a chase through the discovery of strata, fossils and deep time — by way of geologists Nicolaus Steno, James Hutton, Georges Cuvier and Charles Lyell — and through the early history of evolutionary ideas with Charles Darwin and Alfred Russel Wallace. This is well-trodden ground but the stories are clearly and briskly told.

The guts of the book describe the origins of major groups on branches of the vertebrate family tree: tetrapods, birds, mammals, whales, elephants, horses and primates (including humans). Each chapter is as craftily engineered as a computer game. The aim is to entice the reader from tales of quirk and scandal — with the Hungarian baron and his like pressed into service — into the taxonomic and osteological detail. It is pleasing to see credit given to the geologists who made the major discoveries: for instance, to Peter Friend of the University of Cambridge, UK, and his colleagues who first found the Greenland tetrapods long before their many-fingered qualities were dissected by the palaeontologists.

Switek's narrative device generally works well, although occasionally it is in danger of running aground. Detail threatens to overwhelm the chapter

on mammals, for instance. It is the result of an embarrassment of riches, of course, especially in groups such as the whales, for which recent fossil discoveries have illuminated ancestry. Switek, a science writer and blogger, has clearly done a lot of homework. He demonstrates the replacement of old ideas of simple, almost linear evolutionary pathways — such as the classic early picture of horse evolution — with the complexity of modern phylogenetic reconstructions. The wealth of co-existing horse taxa more than 15 million years ago contrasts starkly with the poverty of horse diversity today, for example.

Switek's chapter on human ancestry is one of his best. The pace barely slackens, even as the number of hominin genera mount up. Alas, space seems to have not permitted mention of the awesome *Gigantopithecus*: not a hominin, perhaps, but the tale of its discovery, as 'dragon bones' in an apothecary's shop, would have fitted the book's style nicely. This giant ape, and others nearer our line, perished. Our own species thrives — for now. It is an accident of time and place, as Switek underlines. *Written in Stone* is a fine guide to the four-dimensional tapestry of life — the bony bits of it, at least. ∎

Jan Zalasiewicz *is a senior lecturer in the Department of Geology, University of Leicester, Leicester, LE1 7RH, UK. His latest book is* The Planet in a Pebble. *e-mail: jaz1@leicester.ac.uk*

Rhythm and reason

Nicola Jones witnesses a meeting of dance and ecology.

At a time when many scientists are struggling with how best to communicate with the public, it is refreshing to see the problem approached from the heart. The Canadian dance project *Experiments*, which ran in Vancouver at the end of November and may tour elsewhere, brings together four dancers and four local ecologists to explore synergies between their work.

Choreographer Gail Lotenberg of LINK Dance Foundation aims to convey the human side of research through the performance. She prepared by taping hours of conversations with scientists she met through her husband, Alejandro Frid, a marine biologist at the Vancouver Aquarium Marine Science Centre. Video clips of the interviews — revealing what the scientists do and how they feel about it — are projected onto the back wall of the stage between the dances.

In one piece, the music samples the monologue of marine ecologist Anne Salomon at Simon Fraser University in Burnaby, British Columbia, on the interconnectedness of all things. In another, behavioural ecologist Larry Dill, also at Simon Fraser University, appears on stage in person — complete with Darwinesque beard and white lab coat — observing a dancer's movements just as an ecologist might observe wildlife.

Elements of the dance are inspired by both the process of science and the collaborators' specific research. Lotenberg peers at

Experiments: Where Logic and Emotion Collide
LINK DANCE FOUNDATION, VANCOUVER
25–27 November 2010

the audience with a flashlight to suggest curiosity; a dancer builds a tower of Styrofoam blocks to match a smaller model. The set pieces are beautiful but often rather literal: two dancers dressed in pink and orange interlink to suggest coral and algae in symbiosis; the stage is lit in honeycomb patterns during a piece on bees; dancers in flowing greens and browns appear after Salomon's video describing her work with kelp forests.

Some more-interpretive moments are more revealing. After former ecologist and sculptor Lee Gass tells in a video of his delight at discovering a mathematical representation of the territorial defensive behaviour of hummingbirds, the dancers ponder and reposition cubes placed on stage as if interacting with data points on a graph. And the audience members become part of the 'experiment' when their reaction to a startling event is filmed and played back.

Lotenberg sees many similarities between scientists and artists. Both rely on a pivotal moment of inspiration, she says: "Scientists call it an 'aha' or 'eureka' moment." And the disciplines share rigour: "we repeat, repeat, repeat", scientists to get statistical significance and dancers to rehearse. But these links come across more powerfully through Lotenberg's words than through the dancing.

Mark Winston, a former bee researcher now at the Morris J. Wosk Centre for Dialogue at Simon Fraser University, has worked with Lotenberg before and encouraged the project from its inception. He found taking part in *Experiments* a deeply moving and collaborative experience. Salomon, too, was inspired by the interaction — she says it gave her the idea for a paper on marine conservation. But, she adds, it wasn't like a scientific collaboration in which co-authors all have control: "We were like a bucket of paint to a painter."

Gass says that the project clarified his thinking about the human side of science. He had assumed that in Lotenberg's subtitle, logic referred to science and emotion referred to dance, but later realized that Lotenberg probably meant something else: "Science is where logic and emotion collide." ∎

Nicola Jones *is a freelance writer based in Vancouver, Canada. e-mail: nkjones@gmail.com*

Coral symbiosis comes alive in *Experiments*.

C. RANDLE

CORRESPONDENCE

Non-English papers decrease rankings

In applying a set of standard bibliometric indicators to rank the scientific status of 500 universities worldwide for the 2010 Leiden Ranking, we have discovered that the language of publication has a dramatic and largely underestimated effect on citation-based measurements of research performance.

Publications in non-English-language journals count as part of a country's output, but these generally have a low impact as fewer scientists can read them. This effect is particularly evident in application-oriented fields such as clinical medicine and engineering, and in the social sciences and the humanities.

As clinical medicine represents a considerable part of the scientific output of most nations, the language of publication directly affects the ranking of the university hosting the research (T. N. van Leeuwen *et al. Scientometrics* **51,** 335–346; 2001).

Rankings based on the number of citations per paper and per staff member are responsible for the strikingly low position of many German and French universities, particularly those that include a medical school.

Important ratings by *Times Higher Education*, QS, the Shanghai Academic Ranking of World Universities and the Leiden Ranking, for example, all unfortunately rely on rankings influenced by this language effect.
Ton van Raan, Thed van Leeuwen, Martijn Visser *Leiden University, the Netherlands. vanraan@cwts.leidenuniv.nl*

Citation bubble about to burst?

University rankings would be more informative if they took into account graduates' contributions to a country's international economic competitiveness. Although institutes in the United States and the United Kingdom currently top the university rankings, the most successful technology-based export industries are dominated by northern and central Europe, as well as by Asian city states — the per-capita exports of which are also several times higher.

Rankings are heavily influenced by citations, but these represent little more than symbols. They are comparable to the less-than-worthless collateralized debt obligations that drove the recent financial bubble, and, unlike concrete goods and real exports, they are easy to print and inflate.

Financial deregulation led to short-term incentives for bankers and rating agencies to overvalue their collateralized debt obligations, bringing down entire economies. Likewise, today's academic rankings provide an incentive for professors to maximize citation counts instead of scientific progress (by coincidence, both types of incentive were invented in the United States). We may already be in the middle of a citation bubble — witness how relatively unknown scientists can now collect more citations than the most influential founders of their fields.

It may be easier to collect citation indices than employment histories and other statistics about the impact of graduates on industry, but modifying rankings to take these into account would be worth the effort, particularly for institutes and countries that are currently traded below value.

Note that I write from the country with the most citations per capita and per scientist.
Jürgen Schmidhuber *Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA) and University of Lugano, Switzerland. juergen@idsia.ch*

Measuring impact of research on society

We invite scientists to comment on our proposed tool for rating the impact of research publications on society (www.societalimpact.info). We believe that such ratings should eventually be incorporated into assessments of scientific performance.

This tool has been devised by a task force recruited from different disciplines at the Medical University of Vienna (G. Watts *Br. Med. J.* **338,** b553; 2009). It evaluates factors that include the aim of the published investigation; the extent to which authors attempt to translate their scientific findings into societal action; and the level, status and target group of this translation.

Authors' efforts to convey the significance of their results for daily life are particularly relevant. The level of translation (regional, national or international), its status (preliminary or permanent) and its target group (individuals, population subgroups or the public) are other important quantitative indicators.

There will be challenges in implementing our system. Not least is the problem of measuring the likely societal impact of publications from the basic sciences. There will also be an inherent subjectivity in the assessment, and societal rating may not lend itself to computerized calculation.
Thomas Niederkrotenthaler, Thomas Dorner, Manfred Maier *Medical University of Vienna, Austria. thomas.niederkrotenthaler@ meduniwien.ac.at*

How will growing cities eat?

Food security should not be ignored when assessing the future of our cities (*Nature* **467,** issue 7318; 2010). Urban populations already comprise more than half of humanity and are expanding. This demographic shift will leave fewer farmers to cultivate the food on which cities depend, exacerbating the 20% decline in the number of farmers over the past 40–50 years.

Food for cities will either have to be sourced from remote locations across the globe, or cities will have to incorporate their own food-production facilities by such developments as peri-urban farming.

A scientific question is how cities and the land areas needed to feed them scale in relation to city population density. Scaling would be expected to be positively nonlinear because of the highly variable biological productivity of terrestrial and marine ecosystems that produce this food. In other words, larger cities will sequester proportionately larger and more marginal, low-productivity areas of land in order to be fed. Local urban and peri-urban food production tends to increase the yield per unit area and may partially counteract this trend.

The nonlinear scaling of food provision with city population size contradicts the idea that cities are proportionately more efficient per inhabitant, with respect to their use of infrastructure, carbon emissions and other services (*Nature* **467,** 912–913; 2010). Food security of dense megacities depends on large, low-population land areas that exist elsewhere and yield disproportionately high food surpluses.
John R. Porter *University of Copenhagen, Denmark. jrp@life.ku.dk*
Lisa Deutsch *Stockholm University, Sweden.*
David Dumaresq, Rob Dyball *Australian National University, Australia.*

Frank Fenner

(1914–2010)

A guiding light of the campaign to eradicate smallpox.

On 8 May 1980, the Australian virologist Frank Fenner stood before the World Health Assembly in Geneva, Switzerland — the governing body of the World Health Organization (WHO) — and reported that smallpox, which had plagued mankind for more than 3,500 years, had been eradicated. That moment marked the first time in human history that a disease had been wiped out.

For Fenner, who died on 21 November at his home in Canberra, the announcement also represented the culmination of 35 years of research on poxviruses. This included more than a decade's work on the eradication campaign, first as principal adviser, and latterly as chair of the WHO commission that decided when the virus had been conquered.

Fenner's father was a writer and teacher in the small town of Ballarat, near Melbourne. He passed on his enthusiasm for geography, geology and social history to his family. At his father's urging, Frank studied medicine, graduating from the University of Adelaide in 1939; but he retained wide interests and was working on manuscripts pertaining to ecological and environmental questions up to his death.

With war imminent, Fenner took a diploma in tropical medicine before enlisting. This, he believed, would give him a better chance of receiving the more challenging assignments in disease research and prevention. And so it turned out. During five years of service, primarily in New Guinea, he dealt with all manner of tropical diseases, reducing malaria — then a major drain on Australia's fighting forces — to a trivial problem. He also met his wife, ebullient army nurse Ellen (Bobbie) Roberts, who served as a part-time assistant in his laboratory. She died in 1995.

At 31, Fenner began his research career late. He soon made his presence felt. His first appointment, in 1946, to a fellowship at the Walter and Eliza Hall Institute of Medical Research in Melbourne, proved pivotal. The institute's director was Macfarlane Burnet (few used his first name, 'Frank'), who in 1980 won the Nobel prize for his work on immunology. Burnet tasked Fenner to work on the pathogenesis and epidemiology of ectromelia, the mouse pox virus. Only months before,

Burnet's group had discovered that this virus was in the same family as smallpox, making it a model for studying the human disease.

Fenner's next challenge was myxomatosis or rabbit pox, caused by the myxoma virus. In 1948, massive epidemics of the disease exploded across Australia, killing almost all infected animals. The virus had been introduced into trial sites to evaluate its use in controlling rabbits, but it escaped and millions of rabbits began dying.

In the early 1950s, Fenner, now chair of microbiology at the medical school of the

SYDNEY MORNING HERALD

newly founded Australian National University in Canberra, began comprehensive studies of myxomatosis in the laboratory and the field. As the epidemic peaked, there was an outbreak of human encephalitis, an inflammation of the brain. Media reports blamed the outbreak on the myxoma virus, and vilified the authorities. To allay the panic, Fenner and two colleagues rolled up their sleeves and injected themselves with the virus, to no ill effect. It was a typical, if dramatic, example of his scientific confidence.

In 1967, the WHO began a global programme to eradicate smallpox and I was asked to be its director. That year, more than 10 million cases of the disease and 2 million deaths had occurred in 42 countries. Each year we vaccinated, on average, 300 million people under the programme and contained hundreds to tens of thousands of individual outbreaks.

It was thought that the smallpox virus infected only humans. But greater certainty was needed. If the disease could persist in an animal reservoir — as, say, plague does in wild rodents — then eradication would be impossible. Primates were the prime suspects for such a reservoir. In 1958, a new virus had been discovered that caused a smallpox-like disease in monkeys. In 1969, the WHO convened a group to design field and laboratory studies to investigate the virus's natural behaviour.

Fenner, author of one of the principal virology textbooks, was a member. In 1970, human cases of monkey pox appeared in smallpox-free areas. They looked like typical smallpox. The consultant group reconvened regularly, with Fenner providing the voice of calm and reason. Not until 1979 was it proven that humans infected with monkey pox did not spread the virus easily or quickly, and that cases probably resulted from villagers eating infected rodents. Eradication of smallpox was possible.

As the programme progressed, governments needed convincing that no smallpox remained, and that vaccination and border checks could end. Fenner took the chair of the international commission for certification of eradication, which had the last word on whether the world was free of the virus. They examined the quality of national reports prepared two or more years after the occurrence of a country's last known case and recommended some additional studies.

In December 1979, the committee was satisfied that the disease had been eradicated, and submitted its report to the WHO's director-general. The ready acceptance of the report by the World Health Assembly on that May day in 1980 was a tribute to the wisdom and integrity of Frank Fenner, the programme's distinguished senior mentor and ever-willing consultant throughout the campaign's many difficult years. ∎

Donald A. Henderson *is at the Center for Biosecurity of the University of Pittsburgh Medical Center, and the Johns Hopkins School of Public Health, Baltimore, Maryland 21202, USA.*
e-mail: dahzero@aol.com

NEWS & VIEWS

Simplicity versus complexity

Many scientists now use the power of computer models to advance their subjects. But there is a choice: to simplify complex systems or to include more detail. Modelling the intricate processes of sedimentary geology is a case in point.

LANDSAT 7 / USGS / NASA

The Ganges delta: varying patterns of erosion and silting present challenges for modelling.

In modelling, simplicity isn't simple

CHRIS PAOLA

Before computing became widespread, quantitative modelling usually required simplification — spherical cows[1], if you like — to render complex problems tractable. Fields such as sedimentary geology that are only now becoming quantitative present a dilemma: is there any reason not to skip the first steps and go directly to the most detailed models we can create? Do we really need spherical cows?

I think that the evolution of relatively mature quantitative fields through a stage of creating simplified models was a matter of more than lack of access to large computers. It cultivated an essential counterpart to attention to detail, which is attention to what is truly essential. In general, complexity is ordered and hierarchical[2],

which among other things means that it is not additive. Simple interactions at small scales can produce complex behaviours at larger scales; and complicated small-scale processes can add up to relatively simple large-scale dynamics[2–4]. The discipline imposed by simplification forces one to seek out these effects.

The dynamics and stratigraphy of Earth's surface offer abundant examples of the power of insightful simplification. A generation of reduced-complexity models of landscapes showed how the interplay of simple rules could reproduce major features of tributary-channel networks. Similar approaches, with comparable results, have been applied to such features as stream braiding, sorted sediment patterns, bedforms, beach cusps, coastal morphology and river deltas. In each of these cases, simplified representations of the complex small-scale mechanics of flow and/or sediment motion capture the self-organization processes that create apparently complex patterns.

Moving up in scale, river 'long profiles' reflect the net effect of the complex, nonlinear mechanics of turbulent flow and sediment

transport acting at small scales. Yet for many purposes, long-profile evolution can be represented by relatively simple diffusion models. And emerging evidence suggests that important aspects of large-scale downstream variability in depositional systems, including grain size and channel architecture, can be understood in terms of first-order sediment mass balance.

Traditional training in disciplines such as sedimentary geology understandably emphasizes attention to detail; this can easily carry over into a tendency to think that adding more detail to models is always better. If the goal is simply to produce realistic-looking simulations, this may be correct. But, as Philip England (personal communication) has pointed out, the danger in creating fully detailed models of complex systems is ending up with two things you don't understand — the system you started with, and your model of it.

Simplification is essential if the goal is insight. Models with fewer moving parts are easier to grasp, more clearly connect cause and effect, and are harder to fiddle to match observations. Even when they are only partially correct, simplified models can serve as reference cases[5], a kind of intellectual geoid that lets us subtract first-order dynamics and focus on what is anomalous and idiosyncratic. But we will never find out what simple models can teach us if it becomes an article of faith that the systems we work on are irreducibly complex.

The convergence of high-resolution observational data, peta-scale computing and the obvious complexity of natural systems does make imaginary animals seem a little quaint. But understanding what is really essential requires simplification, which in turn requires imagination — the ability to see beyond what is in front of us — and a willingness to be wrong. Perhaps some day a computer will surprise us by proposing spherical cows. Until then, may they continue to remind us that simulation is not the same as understanding.

Chris Paola *is in the Department of Geology and Geophysics, St Anthony Falls Laboratory, University of Minnesota, Minneapolis, Minnesota 55414, USA.*
e-mail: cpaola@umn.edu

Complexity and the memory of landscape

MIKE LEEDER

The sedimentary record extends over more than 4.5 billion years and is the source for many inferences concerning Earth's history. Over this vast time, Earth has presented many faces, but there is no long-term trend because dynamic processes arise from a multiplicity of variables and time's arrow prevents future recurrence of the totality of past conditions. Yet Earth has certainly 'evolved', if only in the sense that it is always changing in response to the to-and-fro and up-and-down of tectonics, the position of continents and oceans, atmospheric composition, and the solar constant. Orbital parameters and organic evolution are other changing factors. The Pythagoreans had some of this in mind with their idea of the world in a state of constant flux. The concept also lies at the heart of the interdependence of material things recognized by Zen[6] and other philosophies, going back to the Stoics.

In a simple world, solution of the Navier–Stokes equation, which describes fluid flow, would probably suffice to solve problems in what J. J. Thomson[7] memorably referred to as 'out-of-doors physics'. Such disciplines comprise sedimentology, hydrology, geomorphology and other areas of surface environmental dynamics. However, the 'memory effect' makes it impossible for computational schemes to solve the Navier–Stokes equation for naturally evolving fluid flows (chaos theory outwits the weather forecasters in the end). Also, forces at work on the environment result in additional nonlinear effects. Nature is thus inherently complex, refuses to keep to boundary conditions, and is influenced by the pervasive variables of solid geology, climate change and life forms.

The best environmental science recognizes complexity and seeks to write computer code that can at least include influential variables. Take two examples: complex physical feedbacks in sediment transport[8], and models linking climate, vegetation and soil erosion[9]. The former arise when flowing fluid transports solids that affect pressure and viscous forces, and influence fluid accelerations. The latter must deal with the linked complexity of all natural landscapes. The logic of a recent algorithm[9] proceeds as: climate (monthly mean temperature/precipitation) → soil vegetation growth → soil infiltration properties → weather (daily storm events) → run-off → soil erosion potential. It is difficult to see the point in accepting landscape or sediment-routing models that ignore such fundamentals as seasonal water balance and vegetation growth. On the downside, attention to geology and the chemical weathering of rock is still not satisfactorily addressed.

What is the modern geologist to make of all this, for geophysicists contribute little practical help in the field examination of rock? In my opinion, there can be no general theory, only the effects of competing causes. We must try to interpret the rock record by attempting to assign true cause to stratigraphic events that originated in ancient landscapes. For example, what does a pulse of sediment signify? Is there such a thing as stratigraphic equilibrium? What are the timescales of sedimentary response to changes in environmental variables? Can we deconvolve such concatenated variables?

What we really need in 'out-of-doors physics' is a recognition of, on the one hand, the complexity of what I call the ambioscape, being the totality of spatial, environmental and geological variables that define a particular landscape, and, on the other hand, the limits to the explanatory power of self-similarity[10]. ∎

Mike Leeder *is in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.*
e-mail: m.leeder@uea.ac.uk

1. Harte, J. *Consider a Spherical Cow: A Course in Environmental Problem Solving* (Kaufmann, 1985).
2. Werner, B. T. *Science* **284,** 102–104 (1999).
3. Murray, A. B. *Geomorphology* **90,** 178–191 (2007).
4. Bras, R. L., Tucker, G. E. & Teles, V. in *Prediction in Geomorphology* Vol. 135 (eds Wilcock, P. R. & Iverson, R. M.) 63–79 (Am. Geophys. Union, 2003).
5. Power, M. E., Dietrich, W. E. & Sullivan, K. O. in *Experimental Ecology: Issues and Perspectives* (eds Resetarits, W. J. & Bernardo, J.) 113–132 (Oxford Univ. Press, 1998).
6. Watts, A. W. *The Way of Zen* (Pantheon, 1957).
7. Jackson, D. & Launder, B. *Annu. Rev. Fluid Mech.* **39,** 19–35 (2007).
8. La Porta, A., Voth, G. A., Crawford, A. M., Alexander, J. & Bodenschatz, E. *Nature* **409,** 1017–1019 (2001).
9. Kirkby, M. J., Orvine, B. J., Jones, R. J. A., Govers, G. & the PESERA team *Eur. J. Soil Sci.* **59,** 1293–1306 (2008).
10. Werner, B. T. & Kocurek, G. *Geology* **25,** 771–774 (1997).

Bigger and better synthesis

Nature constructs macromolecules with a precision that chemists have struggled to achieve. So a strategy that offers simple routes to large molecules, starting from small templates, could be the next big thing in synthesis. SEE LETTER P.72

CHRISTOPHER HUNTER

Over the past two decades, chemists have exploited developments in polymerization and self-assembly for 'bottom-up' construction techniques, in which small, well-defined molecules bond or interact to make larger objects. But these methods invariably result in mixtures of different products, and so the construction of macromolecular objects with the precision found in biology remains a major challenge for synthetic chemistry. Nature solves this problem by using templates to control both molecular synthesis and supramolecular self-assembly. On page 72 of this issue, O'Sullivan *et al.*[1] report a spectacular strategy for the precise construction of synthetic macromolecules — an approach that they have dubbed 'Vernier templating'.

In 1631, the mathematician Pierre Vernier invented an instrument that could be used to measure lengths and angles with unprecedented precision[2]. The Vernier scale — a sliding, secondary scale that indicates where a measurement lies between two marks on a larger scale — makes it possible to measure minute differences in size using relatively large objects. The principle underlying the instrument is that the graduations in the secondary scale are slightly smaller than those of the main scale, so only the zero point and the last graduation match up with those of the main scale.

The idea to apply Vernier's approach in reverse to synthesize well-defined macromolecular objects from molecular building blocks (Fig. 1a, overleaf) was proposed[3] in 1991. Two different molecules equipped with complementary binding sites will interact to form complexes that grow until all of the binding sites on all of the molecules are occupied. In general, if one molecule has n binding sites and the other has m binding sites, the resulting Vernier assembly will contain a total of $n \times m$ interactions, provided that m is not a multiple of n or vice versa. Just like in the Vernier scale, the assembly process is driven by a mismatch in the size of the template and of the molecules that bind

2011: YEAR OF CHEMISTRY
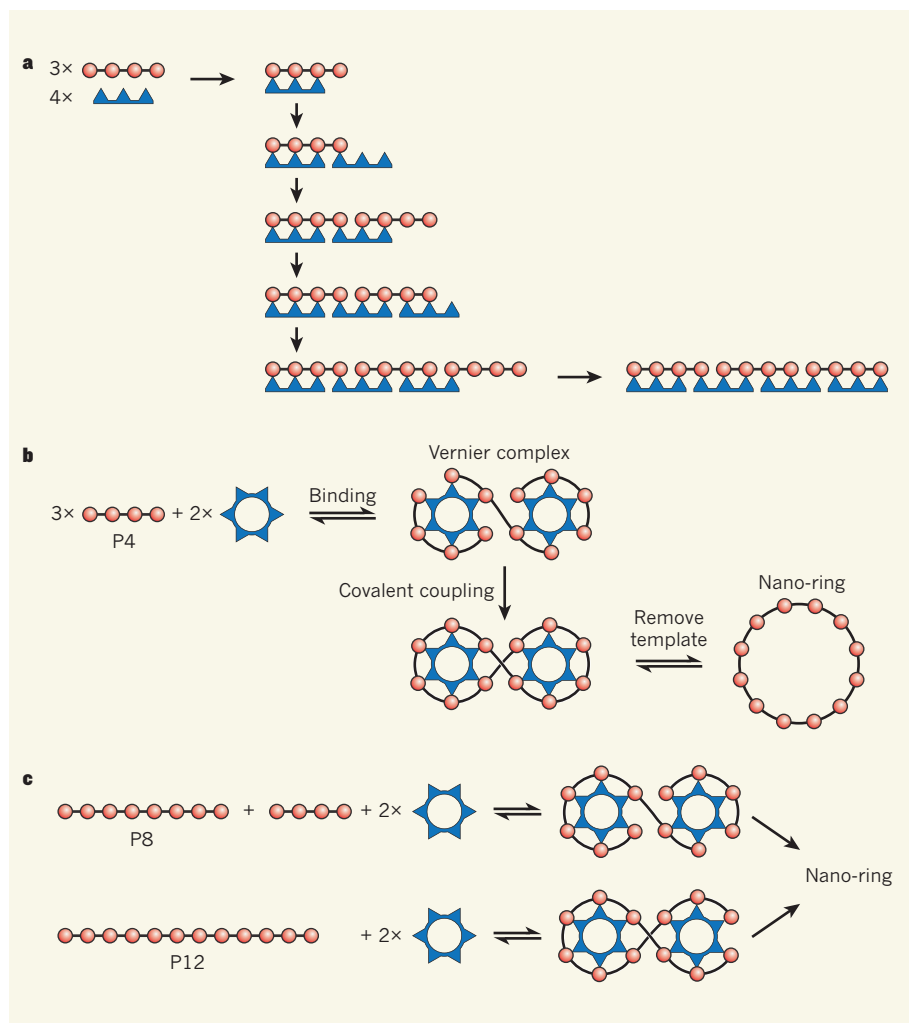Celebrating the central science
nature.com/chemistry2011

Figure 1 | Vernier templates for macromolecular synthesis. a, Vernier assemblies form from molecules that have different numbers of complementary binding sites (here, red circles bind to blue triangles). After the formation of an initial complex, further molecules join the assembly until all of the binding sites on all of the molecules are occupied. **b,** O'Sullivan and colleagues[1] used this concept to assemble nano-rings from tetrameric subunits containing four binding sites (porphyrin groups), starting with a ring-shaped template that has six complementary binding sites. Three subunits (P4) assemble with two templates, forming a figure-of-eight complex. Covalent coupling of the subunits in the complexes, followed by removal of the template, allows the preparation of nano-rings containing 12 linked porphyrins. **c,** Octamers (P8) or dodecamers (P12) of porphyrins can form during the reaction. These also bind to the template, either alone (dodecamers) or in combination with a tetramer (octamers). After covalent coupling and template removal, the resulting complexes form the same nano-rings as before. In **b** and **c,** the figure-of-eight complexes may form as different isomers, all of which lead to the same nano-ring; only one isomer of each complex is shown.

to it. Vernier templating thus provides an attractive strategy for making macromolecules from much smaller templates with molecular precision.

This approach has previously been used in molecular self-assembly processes[4], but because the resulting Vernier products are held together with reversible, non-covalent bonds, they are fragile. What's more, partially assembled structures form alongside the desired, fully assembled product (Fig. 1a). Connecting all of the components of the assembly with covalent bonds is a possible strategy for trapping the product in a more robust form; however, O'Sullivan and colleagues' approach[1]

takes things one step further with direct Vernier templating of a cyclic covalent product (a molecular ring). This strategy provides a particularly efficient synthesis, because once the molecules have been sealed into a closed ring, there are no dangling end groups that could react further to produce polymeric mixtures.

In their study, the authors used a ring-shaped template to construct large cyclic molecules from linear subunits (Fig. 1b; see Fig. 2 on page 73 for the chemical structures of the template, and Fig. 4 on page 74 for the structure of the product). These subunits consisted of four porphyrin groups, connected together in a linear chain. In the absence of a

template, the authors found that the formation of covalent bonds between the porphyrin tetramers (referred to as P4 molecules) yielded an ill-defined mixture of polymeric products. But in the presence of a template containing six sites to which porphyrin groups can bind, the authors observed the formation of Vernier assemblies: complexes in which three P4 molecules wrap around two templates to form figure-of-eight structures. When the authors formed covalent bonds between the P4 molecules in the complex and then removed the template, they obtained a cyclic product — a nano-ring containing 12 porphyrins. This is one of the biggest molecules ever synthesized as a pure compound.

The beauty of O'Sullivan and colleagues' strategy is that efficient formation of the Vernier assembly is not required. In the reaction that covalently connects P4 subunits, any P4 molecules that aren't part of a Vernier assembly will connect together to make porphyrin octamers (P8 molecules). Because P8 has eight porphyrins, it will bind even more strongly to templates than P4, and will contribute to figure-of-eight complexes in which one P8 and one P4 wrap around two templates; subsequent covalent trapping of these complexes will again yield the desired nano-rings (Fig. 1c). Similarly, if any P8 escapes this pathway, then it will most probably react with another P4 molecule to form a porphyrin dodecamer, P12. This compound will bind yet more strongly to templates, and covalent trapping of the resulting complex will again yield a nano-ring consisting of 12 porphyrins (Fig. 1c). So, even under conditions in which the primary Vernier complex is never significantly populated, the template mechanism funnels any oligomeric intermediates — the side-products of untemplated reactions — to the desired product and prevents them from polymerizing. This built-in correction mechanism results in a remarkably efficient synthesis.

Not only does O'Sullivan and colleagues' work[1] represent a great advance in macromolecular synthesis, but the compounds thus obtained are also remarkable. The structure of the nano-ring is reminiscent of the chlorophyll arrays used by plants in photosynthesis for light-harvesting and energy conversion[5]. Photochemical and electrochemical studies of the nano-rings are likely to identify unusual properties that will provide a challenge for theory to explain.

One of the exciting things about the authors' report[1] is that their research group has already assembled a range of different molecular templates, as part of previous studies[6] on conventional template-directed synthesis. The stage is now set for the synthesis of a range of new macromolecules of unprecedented dimensions. For example, the authors previously used a dodecameric template (one that has 12 porphyrin-binding sites) to make the nano-ring reported in the current work[1], but less

efficiently. If this template was combined with a porphyrin pentamer in a Vernier synthesis, then the product would be a ring of 60 porphyrin units in circumference, with a molecular mass of 48 kilodaltons and a cross-section of 25 nanometres. This is the size of an average protein or nanoparticle, and is well beyond the realms of conventional chemical synthesis. A simple and efficient synthetic route to such large molecules might provide all kinds of new materials for nanotechnology. ∎

Christopher Hunter *is in the Department of Chemistry, University of Sheffield, Sheffield S3 7HF, UK.*
e-mail: c.hunter@shef.ac.uk

1. O'Sullivan, M. C. *et al. Nature* **469**, 72–75 (2011).
2. Vernier, P. *La Construction, l'usage, et les propriétés du quadrant nouveau de mathématiques* (1631).
3. Lindsey, J. S. *N. J. Chem.* **15**, 153–180 (1991).
4. Hunter, C. A. & Tomas, S. *J. Am. Chem. Soc.* **128**, 8975–8979 (2006).
5. McDermott, G. *et al. Nature* **374**, 517–521 (1995).
6. Hoffmann, M., Wilson, C. J., Odell, B. & Anderson, H. L. *Angew. Chem. Int. Edn* **46**, 3122–3125 (2007).

EVOLUTIONARY BIOLOGY

Catfish mimics

Mutualism can be a double-edged sword if the animals concerned also compete for food. This may explain the discovery that catfish mimics in the Amazon rarely engage in mimicry with related species. SEE LETTER P.84

JAMES MALLET & KANCHON DASMAHAPATRA

Müllerian mimicry in the warning colours of unpalatable butterflies has been well known since its discovery in the nineteenth century[1,2]. It is beneficial because noxious prey that share warning colours also share the cost of educating predators about their unpalatability[3]. Müllerian mimicry is now known to occur in many other animal groups, including millipedes[4] and poison dart frogs[5].

Below the surface of the Amazon, on whose banks Henry Walter Bates made the first discovery of butterfly mimicry, a new example of Müllerian mimicry has just been discovered by Alexandrou *et al.* (page 84 of this issue)[6]. The authors studied armoured catfish in the subfamily Corydoradinae, which swim in multi-species shoals and are defended by retractable venomous spines and bony plates. Across the Amazon basin, it turns out that up to three unrelated species with strikingly similar colour patterns may coexist in any one locality. The bold, mimetic patterns of these fishes (Fig. 1) undoubtedly serve to warn predators away, as recently demonstrated in an unrelated African catfish genus[7].

Bates's discovery of mimicry stemmed, in part, from his finding that local colour patterns converged, but more importantly from the fact that butterflies "are found all to change their hues and markings together, as if by the touch of an enchanter's wand, at every few hundred miles" (H. W. Bates, quoted in ref. 2). This striking and repeated phylogenetic pattern of convergence across the Amazon clinched the evolutionary hypothesis of mimicry: species differ in colour from related forms in distant areas, while at the same time mimicking unrelated species locally. A spatial pattern of local convergence and geographical divergence

is quite general in other systems of Müllerian mimicry[4,5], as well as in the catfish studied by Alexandrou and colleagues[6].

However, these patterns are even stranger than they first appear. Mimicry explains why species converge, but cannot explain why species also diverge into multiple mimicry 'rings' in many Müllerian mimetic systems. Diversity is often local: in armoured catfish there may be up to six mimicry rings in some regions[6] (although they rarely swim together), and in *Heliconius* butterflies it almost seems as if species undergo adaptive radiation[8] into different mimicry 'niches'. Why should new mimicry rings ever evolve, given that mimicry itself should prevent divergence? Slow neutral change of pattern in geographically distant

regions might explain some of the diversity. However, although distance can prevent the swamping effect of gene flow, it cannot drive divergence directly.

Consider the diversity of mimicry in its overall ecological context. Müllerian mimicry is only one of many interactions among species. It provides an example of a positive interaction, or mutualism. To explain species coexistence, we must understand how this mutualism combines with negative interactions, such as competition. Competitively dominant species typically exclude inferior competitors. However, recent theory shows that if inferior competitors develop mutualisms with dominant species, this can allow them to invade and coexist, permitting higher diversity than without mutualism[9]. In the ithomiine butterflies, the classic group of Müllerian mimics[1,2], this may explain why some closely related species that use similar resources are also close mimics[10].

Now imagine a superior competitor associated with multiple, mimetic mutualists. This dominant species reduces predation and enables mimics and inferior competitors to coexist, inevitably leading to some cost to itself[9]. In the first ever mathematical evolutionary theory, Müller[2] proved that the relative advantage of mutualistic mimicry to each species is roughly proportional to the inverse square of relative abundance: in other words, rarer mimics benefit very much more from Müllerian mimicry than commoner mimics. A more common species could suffer a considerable amount of resource competition if it is helping multiple co-mimics, while benefiting little from mimicry itself.

Therefore, a superior competitor species that escapes its ancestral mimicry could rid itself of some competitors. It could then spread at the

Figure 1 | Warning signals. The catfish *Corydoras haraldschultzi*, with its bold markings, is one of the Amazon mimetic species studied by Alexandrou and colleagues[6].

MARTIN I. TAYLOR

expense of a less fortunate related form that continues to help competitors through mimicry. This idea cannot explain how the required speciation and mimicry divergence was initiated; but given that it did (and does) happen, it would have the consequence that dominant species would escape pesky competitors. This would stabilize the diversification of the mimicry we observe in nature.

The Amazon catfish[6] seem to corroborate the escape idea. More than 90% of mimics differ from co-mimics in snout length and stable nitrogen isotope content, both indicators of diet: these catfish seem not to compete for resources with co-mimics. Typically, co-mimics are also unrelated. Closely related catfish, by contrast, are usually similar in terms of snout morphology and diet, and tend not to be co-mimics. Escape from mutualistic mimicry could similarly help to explain the 'adaptive radiation' of *Heliconius* butterfly mimicry[8].

Nature is, however, full of contrasts. In unrelated venomous catfishes from Africa's Lake Tanganyika, and in ithomiine butterflies, mimics are often closely related and occupy similar habitats[7,10]. Escape from potentially onerous mutualism has not always occurred. Perhaps resource competition is ameliorated in some other way; or perhaps the dominant species in each ring is unlucky and has

simply been unable to escape the resource pressure exerted by the mimetic mutualists that it unwittingly saves from extinction. Additional theoretical and experimental studies are needed.

Meanwhile, Alexandrou and colleagues' findings[6] in catfish show that mimicry is still contributing fundamental ideas to ecological and evolutionary biology, 150 years after its discovery. ∎

James Mallet *and* **Kanchon Dasmahapatra** *are in the Galton Laboratory, University College London, 4 Stephenson Way, London NW1 2HE, UK.* e-mail: j.mallet@ucl.ac.uk

1. Bates, H. W. *Trans. Linn. Soc. Lond.* **23,** 495–566 (1862).
2. Müller, F. *Proc. Entomol. Soc. Lond.* **1879,** xx–xxix (1879).
3. Ruxton, G. D., Sherratt, T. N. & Speed, M. P. *Avoiding Attack: The Evolutionary Ecology of Crypsis, Warning Signals, and Mimicry* (Oxford Univ. Press, 2004).
4. Marek, P. E. & Bond, J. E. *Proc. Natl Acad. Sci. USA* **106,** 9755–9760 (2009).
5. Symula, R., Schulte, R. & Summers, K. *Proc. R. Soc. Lond. B* **268,** 2415–2421 (2001).
6. Alexandrou, M. A. *et al. Nature* **469,** 84–88 (2011).
7. Wright, J. J. *Evolution* doi:10.1111/j.1558-5646.2010.01149.x (2010).
8. Turner, J. R. G. *Zool. J. Linn. Soc.* **58,** 297–308 (1976).
9. Gross, K. *Ecol. Lett.* **11,** 929–936 (2008).
10. Elias, M., Gompert, Z., Jiggins, C. & Willmott, K. *PLoS Biol.* **6,** e300 (2008).

Toxic Cambrian oceans

Isotopic evidence from carbon and sulphur points to the spread of anoxia and toxic sulphide as the chief culprits in at least one of a series of crises for marine ecosystems during the nascent stages of early animal evolution. SEE LETTER P.80

GRAHAM SHIELDS-ZHOU

By 520 million years ago, the oceans of the world were teeming with diverse forms of animal life addicted to the newly abundant, life-giving oxygen in the surface environment. Arthropods, many — like the trilobites — now extinct, ruled the seas, overshadowing the relatively few extant members of our own animal group, the chordates. But the course of this 'Cambrian explosion' of life did not run smoothly, and on page 80 of this issue Gill *et al.*[1] illuminate a likely cause.

For almost half a century it has been recognized that many of these early animal groups suffered a series of crises, or 'turnovers'. Not only did many trilobite groups diversify, only to become extinct shortly after, but they also seem to have been replaced abruptly by exotic groups taking advantage of the newly emptied ecosystem niches. The spread of anoxic sea water has long been implicated in the apparent volatility of early marine ecosystems, but evidence for

that possibility has been sparse. Gill *et al.* now present geochemical analyses and modelling of four geochemical tracers (carbon and sulphur isotopes, iron speciation and molybdenum content) from six Cambrian-age stratigraphic sections from around the world. In doing so, their work adds considerably to our appreciation of how early marine ecosystems were repeatedly devastated by highly variable levels of oxygen and sulphide.

Allison R. (Pete) Palmer was one of the first to recognize that trilobites and other Cambrian fauna suffered a series of extinctions, some of them major, between about 520 million and 490 million years ago. In 1965, he coined the term 'biomeres' for such intervals of apparent stasis followed abruptly by extinction and replacement[2]. Palmer and others put these extinctions down to changing oceanographic conditions that caused anoxic and/or cooler sea water to invade the shallow, coastal-shelf environment. But this hypothesis remained largely in the realm of speculation.

Although it wasn't obvious at the time, a big step forward came in 1993 with the report[3] of a single carbon-isotope 'excursion' — an anomalously high $^{13}C/^{12}C$ ratio ($\delta^{13}C$ value) — from a limestone sample overlying the base of one of Palmer's biomeres. The beginning of the interval concerned is now dated to 499 million years ago, and it probably lasted for 2 million to 4 million years. It was characterized by biological turnover around the world, and the isotope excursion — known as the Steptoean Positive Carbon Isotope Excursion (SPICE) — has subsequently been shown to be global in extent. Increases in seawater $\delta^{13}C$ are generally believed to be the result of episodes of elevated organic-carbon burial, which can be caused by heightened ocean productivity and/or preservation of organic matter, both of which can be due to anoxia.

Sulphur in the marine environment undergoes analogous isotopic fractionation to carbon, with increased $^{34}S/^{32}S$ ratios in sea water resulting from enhanced burial of reduced sulphur in the form of pyrite (FeS_2). Gill *et al.*[1] report that parallel $\delta^{34}S$ and $\delta^{13}C$ excursions in both sulphate and sulphide $\delta^{34}S$ occurred widely across the SPICE event. This shows convincingly that organic carbon burial was accompanied by an equally spectacular rise in pyrite burial, first hinted at in an earlier, pioneering study[4] of seawater $\delta^{34}S$. The removal of reduced carbon and sulphur into the sediments would have led to a surge in oxygen levels in the atmosphere, and so would hardly seem consistent with the spread of anoxia. But Gill *et al.*[1] marshal enough geochemical evidence, and point to a modern analogy, to convince us that this is precisely what occurred.

In the modern marine environment, parallel burial of organic matter and pyrite chiefly occurs beneath anoxic water columns that are also rich in hydrogen sulphide. Under these special conditions, anoxia restricts organic decay, while allowing soluble Fe^{2+} to build up in the water. Bacterial sulphate reduction supplies the sulphide, which, on binding with the iron, precipitates out as pyrite. Gill *et al.* suggest that the Cambrian seas during the SPICE excursion must have been rather like the modern Black Sea, with oxygen-rich waters at the surface and sulphidic waters at depth, only covering a much greater portion of the oceans and ocean margins. The SPICE interval saw a considerable change in sea level, from a high stand at the outset to lower levels. The authors propose that the result was that anoxic conditions prevailed widely in coastal areas, the same places that are normally most conducive to life.

Other predicted consequences of widespread aqueous sulphide — depletion of the molybdenum reservoir, low carbon/sulphur ratios in shales, and reduced sulphate–sulphide isotopic fractionation — are confirmed in the study[1]. They all help to support the connection between biological extinction and anoxia during the Cambrian.

Using a 'box model' of the global carbon and sulphur cycles, Gill et al.[1] also estimate that the ocean sulphate reservoir was then a small fraction of its current size. That perhaps explains why, compared with today, the sulphur-isotope system during the Cambrian seems to have been exceptionally sensitive to perturbations. However, there are inconsistencies here that require further attention. Despite the estimate of a small sulphate reservoir, the authors found puzzlingly high concentrations of sulphate in many of their rock samples, which remain to be explained. As Gill et al. point out, the sulphur-isotope trends and values also seem to differ between localities, suggesting that not all of the sulphur-isotope values are representative of the global ocean. Although spatial heterogeneity of δ^{34}S is a key prediction of a low-sulphate ocean, such variability makes the authors' size estimates for the marine sulphate reservoir uncertain.

Nevertheless, these new geochemical data[1] are remarkable, in that they pass all geochemical tests of Palmer's anoxia–extinction hypothesis. The distribution of ocean anoxia, and in particular the periodic expansion of toxic sulphide-rich waters, was possibly one key factor that determined how animals evolved on Earth. ■

Graham Shields-Zhou *is in the Department of Earth Sciences, University College London, London WC1E 6BT, UK.*
e-mail: g.shields@ucl.ac.uk

1. Gill, B. C. *et al. Nature* **469,** 80–83 (2011).
2. Palmer, A. R. *J. Paleontol.* **58,** 599–611 (1984).
3. Brasier, M. D. *Geol. Soc. Lond. Spec. Publ.* **70,** 341–350 (1993).
4. Kampschulte, A. & Strauss, H. *Chem. Geol.* **204,** 255–286 (2004).

QUANTUM PHOTONICS

Entangled photons on a chip

Using photonic chips to control single photons in waveguides is a promising route to technologies based on the photons' quantum properties. The ability to measure entanglement on such chips is a key step in that direction.

MIRKO LOBINO & JEREMY L. O'BRIEN

The counterintuitive properties of quantum entanglement have been a topic of intense (meta)physical debate since the idea's inception at the start of the last century. Einstein famously thought that the "spooky action at a distance" arising from entanglement meant that the theory of quantum mechanics was incomplete; it has subsequently proved to be the most successful theory ever developed in terms of its predictive power. However, entanglement and other weird properties of quantum mechanics are not just a scientific curiosity. Researchers around the world are trying to harness them to gain unprecedented power and functionality in information processing, secure communication and precision measurement. Writing in *Physical Review Letters*, Sansoni et al.[1] demonstrate a particularly attractive type of measurement of quantum entanglement between photons that brings these quantum technologies a step closer.

Among the physical systems being used to develop quantum technologies, photons are particularly appealing because of their low noise (information preserving), high-speed transmission and ease of manipulation at the single-photon level[2]. They are indispensable for quantum communication[3] and quantum metrology[4], and of the various approaches to the longer-term goal of quantum computing[5], photonics is a leading candidate[6].

The remarkable properties of quantum entanglement are beautifully exhibited by two 'polarization entangled' photons. Polarization is the direction in which light's electric field oscillates. When the polarization of one of the photons is measured, the outcome is completely uncertain — half of the time it will be horizontal and the other half vertical, or half at +45° and half at −45°, depending on the direction in which it is measured. However, the outcome of the same measurement on the second photon is precisely known: it will either give the same result as for the first photon, in which case the measurements are perfectly correlated; or it will always be the opposite (that is, horizontal for vertical, +45° for −45°, and so on), in which case they are perfectly anti-correlated. (Whether they are perfectly correlated or perfectly anti-correlated depends on the exact way in which the photons are entangled.)

By itself, that's not so surprising. Such correlations are common in our everyday experience. For example, the presence of your keys in your pocket is perfectly correlated with them not being on the hook at home. But when these measurements are repeated on many pairs of such identically prepared entangled photons, no matter at what angles the polarization is measured (horizontal, vertical, diagonal, and so on), the same perfect correlations are seen — and this is clearly at odds with our intuitive understanding of how the world works.

Sansoni and colleagues[1] now demonstrate the measurement of such polarization entanglement of two photons using an integrated waveguide device that incorporates waveguide channels on a chip. The channels guide light in the same way as the optical fibres used in telecommunications (by means of total internal
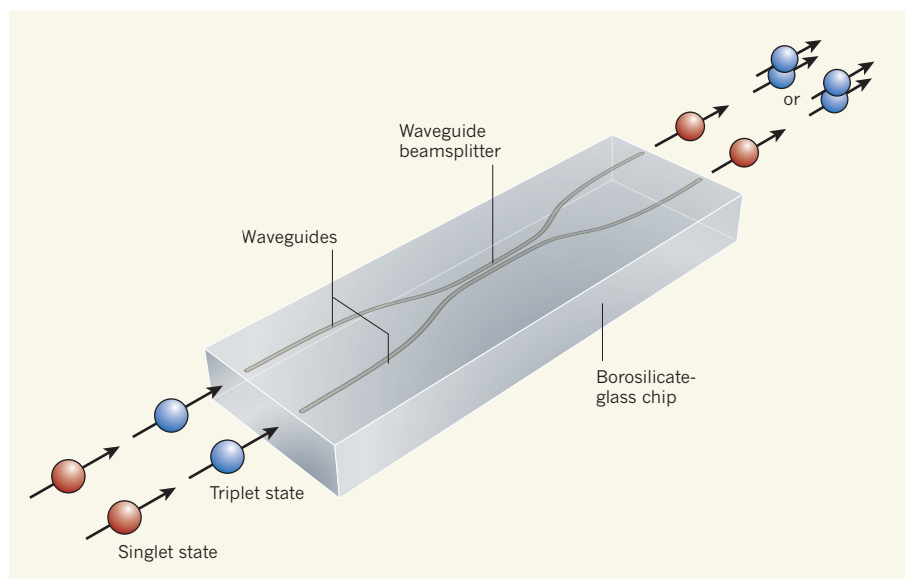

Figure 1 | Measuring entanglement on a chip[1]. A pair of photons is launched into a waveguide beamsplitter on a borosilicate-glass chip. By measuring whether the photons exit in separate waveguides or together in the same waveguide, the nature of the photons' entangled input state — singlet or triplet — can be determined.

reflection in the waveguides), and the device allows the authors to identify the type of polarization entanglement carried by the pair of photons (Fig. 1).

A major obstacle to quantum technologies based on photons, and indeed to the pursuit of fundamental quantum science with photons, has been the construction of photonic quantum circuits from bulk optical elements (such as centimetre-sized mirrors and beamsplitters) that are bolted to large, vibration-stabilized optical tables, with the photons propagating in free space. This approach is unwieldy, unscalable and limited in performance. Recently, quantum circuits that incorporate silica-glass waveguides on silicon chips have been used to overcome these problems[7], following in the footsteps of devices developed for optical telecommunications. This alternative technique has been used to demonstrate small-scale quantum information processing algorithms[8], quantum metrology with up to four entangled photons[9], and a quantum 'walk' of two correlated photons[10].

However, in all of these demonstrations, the quantum information carried by the photons has been path encoded — that is, a photon in one waveguide is used to encode a logical bit 0, with a photon in a second waveguide encoding a logical bit 1. This path encoding is incompatible with the commonly used and very convenient polarization encoding — for example, horizontal photon polarization for a 0 and vertical polarization for a 1. Using polarization encoding in waveguides is not straightforward because the speed with which these two polarizations of light propagate is typically slightly different owing to an effect called birefringence, which is caused by the geometry and/or the materials from which the waveguides are made. This birefringence changes the polarization state of photons propagating in the waveguide, from diagonal to circular for example, and can ultimately depolarize the photons, limiting their capacity to interfere and leading to a loss of information.

An alternative approach to constructing waveguides, used by Sansoni *et al.*[1], involves tightly focusing a powerful pulsed laser beam onto a material such as glass and so permanently changing the material's properties to create a waveguide. By scanning the position of the focused spot around the sample, waveguides can be directly written in all three dimensions of the sample. Quantum photonic circuits with circular waveguides[11] have been made in this way — suggesting that the waveguides would exhibit low birefringence and be able to support polarization encoding. In their study, Sansoni *et al.*[1] demonstrate that the approach can indeed lead to waveguides that support polarization encoding.

They used the technique to fabricate two waveguides on a glass chip and bring them into close proximity, so making a device — termed a directional coupler — that acts like a beamsplitter: it reflects half of the incident light and transmits the other half. When an entangled pair of photons was launched into the two waveguides, the symmetry of the entanglement injected into the device — technically called 'singlet' or 'triplet' states — could be determined. This is because the singlet state 'anti-bunches' at the coupler such that the photons come out of each output waveguide separately, whereas the triplet states 'bunch' such that both photons come out in the same waveguide (Fig. 1).

This polarization-compatible approach to on-chip integrated quantum photonics holds great potential for being combined with the standard methods for generating and measuring photonic entanglement. As the authors[1] point out, it should make the implementation of sophisticated multi-photon quantum circuits such as entanglement filters[12], and even more advanced circuits, much more straightforward. However, integrated devices that can perform polarization rotations in analogy with their bulk optical counterparts — 'integrated waveplates' — will ultimately be crucial, as will a path to miniaturizing these circuits further to make large

quantum photonic circuits possible. As with all approaches to quantum photonic technologies, single-photon sources and detectors remain an important requirement, together with their integration with quantum circuits[2]. ∎

Mirko Lobino *and* **Jeremy L. O'Brien** *are at the Centre for Quantum Photonics, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, UK.*
e-mail: jeremy.obrien@bristol.ac.uk

1. Sansoni, L. *et al. Phys. Rev. Lett.* **105,** 200503 (2010).
2. O'Brien, J. L., Furusawa, A. & Vučković, J. *Nature Photon.* **3,** 687–695 (2009).
3. Gisin, N. & Thew, R. *Nature Photon.* **1,** 165–171 (2007).
4. Nagata, T., Okamoto, R., O'Brien, J. L., Sasaki, K. & Takeuchi, S. *Science* **316,** 726–729 (2007).
5. Ladd, T. D. *et al. Nature* **464,** 45–53 (2010).
6. O'Brien, J. L. *Science* **318,** 1567–1570 (2007).
7. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. *Science* **320,** 646–649 (2008).
8. Politi, A., Matthews, J. C. F. & O'Brien, J. L. *Science* **325,** 1221 (2009).
9. Matthews, J. C. F., Politi, A., Stefanov, A. & O'Brien, J. L. *Nature Photon.* **3,** 346–350 (2009).
10. Peruzzo, A. *et al. Science* **329,** 1500–1503 (2010).
11. Marshall, G. D. *et al. Opt. Express* **17,** 12546 (2009).
12. Okamoto, R. *et al. Science* **323,** 483–485 (2009).

ALZHEIMER'S DISEASE

Recollection of lost memories

With age comes wisdom, or so they say. The reality is that, with age, the ability to store memories declines. One way of tackling this problem might be to raise neuronal levels of the signalling molecule EphB2. SEE ARTICLE P.47

ROBERT C. MALENKA & ROBERTO MALINOW

Where did I put those keys? What did I have for dinner last night? Cognition — most notably, the ability to store memories — inevitably declines with age. What's more, for an increasing proportion of individuals, this decline progresses aggressively to the point that they cannot care for themselves. Alzheimer's disease is the leading cause of such dementia in the elderly, affecting almost 50% of people over the age of 85. But despite considerable progress in understanding the biology of this disease, an effective treatment remains elusive. On page 47 of this issue, Cissé *et al.*[1] provide compelling evidence that manipulation of a specific membrane protein — the receptor tyrosine kinase EphB2 — in a mouse model of Alzheimer's disease can reverse the characteristic memory deficits and so may make for a promising therapeutic strategy.

The leading hypothesis for the cause of Alzheimer's disease — based initially on human genetic findings, and supported by many cell-biological, animal-model and human studies — is chronically high brain levels of a peptide fragment termed Aβ. Indeed, mutations in the enzymes that generate Aβ or in the Aβ precursor protein, which lead to increased Aβ levels, are associated with early-onset Alzheimer's[2,3].

Mice genetically engineered to express these same mutations develop cognitive deficits as they age[2,3]. In normal mice, meanwhile, raising neuronal Aβ levels causes a loss of synaptic junctions between these cells that correlates well with the degree of dementia in humans. Furthermore, Aβ can now be imaged non-invasively in the human brain, and brain images of patients with Alzheimer's disease show Aβ accumulation, the extent of which correlates with memory decline[4]. Thus, there is great motivation to determine how Aβ accumulation leads to memory impairment.

Learning and memory are thought to require

long-term potentiation (LTP) of transmission at synapses — a form of plasticity that occurs prominently in the hippocampus region of the brain. The hippocampus is not only required for memory formation, but is also affected early on during Alzheimer's disease[2,3,5]. It is perhaps not surprising, therefore, that hippocampal LTP is also impaired in mouse models of the disease and after Aβ application[2,3]. But the detailed molecular mechanisms that underlie the impairments in LTP and memory in models of Alzheimer's disease are unknown. More importantly, reversing these impairments has proved very difficult.

To make headway on these problems, Cissé et al.[1] examine the mechanisms behind the loss of a crucial synaptic protein complex, the NMDA receptor (NMDAR). This receptor is required for triggering LTP and for hippocampus-dependent memory formation[5], so it makes sense that its loss and malfunction would contribute to the symptoms of Alzheimer's disease. Indeed, previous studies have found that Aβ both reduces the synaptic function of NMDARs[6] and triggers their internalization from the cell surface[7].

Cissé and colleagues focus on EphB2. This protein interacts with NMDARs[8], and its deficiency reduces LTP[9,10]. The authors find that Aβ binds to EphB2, decreasing its levels. What's more, the effects of reducing EphB2 levels in the dentate gyrus — the input region of the hippocampus — in the normal mouse brain mimic the reduced synaptic NMDARs and LTP that occur in an Alzheimer's disease mouse model.

These findings lay the foundation for the key question: can virus-mediated expression of EphB2 in the dentate gyrus of an Alzheimer's mouse model overcome the associated synaptic and memory deficits? Remarkably, Cissé et al. report that this manipulation 'cures' the mice, with both NMDAR-mediated synaptic responses and LTP returning to normal levels. Of greatest clinical relevance, however, is the authors' finding that EphB2 expression allows the Alzheimer's mice to learn and remember normally in three different behavioural tasks that test hippocampus-dependent memory.

Cissé and co-workers' observations provide compelling evidence that EphB2 may be a valuable target for the treatment of Alzheimer's disease. The authors point out several ways by which treatment might be achieved, such as interfering with the binding of Aβ to EphB2, decreasing EphB2 degradation and increasing EphB2 expression. But before researchers in academia and the drug industry vigorously pursue this novel therapeutic mechanism, some cautionary notes are worth mentioning.

Given the complexity of this challenging subject, it is essential that these key findings be replicated. Moreover, which of the Alzheimer's disease models, if any, accurately reflects the human condition is debatable. As a result,

whether EphB2 expression rectifies synaptic and memory deficits in other models of the disease should also be tested.

Alzheimer's disease affects several brain regions. So it is surprising that, in Cissé and colleagues' mice, expression of EphB2 in the dentate gyrus alone could completely overcome the memory deficits. As the authors suggest, improving the function of a subset of neurons may be sufficient to improve the performance of the larger networks in which they function. Nonetheless, it must be determined whether EphB2 can remedy the Aβ-induced deficits in other brain areas.

Finally, a key challenge is determining the time point during the progression of Alzheimer's disease at which administering a certain treatment would still be effective. Cissé et al. expressed EphB2 before any of the classic neuropathological features of Alzheimer's disease occurred. It should be determined whether expression of this receptor at later time points is equally effective.

Despite these caveats, this work opens up a new avenue of investigation into the pathogenic mechanisms of Alzheimer's disease, and points to a previously unknown mechanism that can be targeted for therapeutic purposes. It also offers hope to a research field that has

recently suffered several highly publicized failed clinical trials. These findings[1] should provide renewed energy and optimism that will hopefully lead to new drugs in time for many of us to take before we develop this devastating disorder. ∎

Robert C. Malenka is in the Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, California 94305, USA.
Roberto Malinow is in the Departments of Neuroscience and Biology, University of California, San Diego, La Jolla, California 92093, USA.
e-mails: malenka@stanford.edu; rmalinow@ucsd.edu

1. Cissé, M. et al. Nature **469,** 47–52 (2011).
2. Tanzi, R. E. & Bertram, L. Cell **120,** 545–555 (2005).
3. Walsh, D. M. & Selkoe, D. J. Neuron **44,** 181–193 (2004).
4. Jack, C. R. Jr et al. Brain **133,** 3336–3348 (2010).
5. Malenka, R. C. & Bear, M. F. Neuron **44,** 5–21 (2004).
6. Kamenetz, F. et al. Neuron **37,** 925–937 (2003).
7. Snyder, E. M. et al. Nature Neurosci. **8,** 1051–1058 (2005).
8. Dalva, M. B. et al. Cell **103,** 945–956 (2000).
9. Henderson, J. T. et al. Neuron **32,** 1041–1056 (2001).
10. Grunwald, I. C. et al. Neuron **32,** 1027–1040 (2001).

MOLECULAR COMPUTING

DNA as a logic operator

Computers use transistor–based logic gates as the basis of their functions, but molecular logic gates would make them much faster. A report of DNA–based logic gates could be a first step towards molecular computing.

THOMAS CARELL

The DNA molecule, which encodes the complexity of all life forms on Earth, is arguably the most powerful medium known for data storage and processing. Yet, to date, DNA has found little use in computing applications. Reporting in *Angewandte Chemie*, Park et al.[1] describe a way in which the molecule of life can be turned into devices able to perform logical operations.

The basis of contemporary computers is Boolean logic, a system in which a 'gate' unit converts input signals into a defined output. The gates have names such as AND, OR and YES, to describe the kind of output that is generated in response to different inputs. Because the output of one gate can be the input of the next gate, hierarchical Boolean networks can be established to compute answers to complex questions. In silicon-based computers, gates are implemented electronically using transistors, which are assembled into integrated circuits. Park et al.[1], however, have constructed

logic gates from DNA. They have used DNA replication as the output signal and built control units into the gates to switch replication on or off depending on specific inputs.

DNA replication is a complex process in which the DNA polymerase enzyme makes duplicates of each strand of a template DNA, starting from two smaller single strands that act as 'primers' (Fig. 1). These primers must bind tightly to complementary sequences at the ends of the template DNA strands to allow replication. Park et al.[1] used modified DNA primers as the basis of their logic gates — primers in which a single DNA base had been changed, so that the substitute base could no longer pair up with the complementary base in the DNA template. The presence of such a mismatch prevents the primer from initiating DNA replication. So far, so good. But the
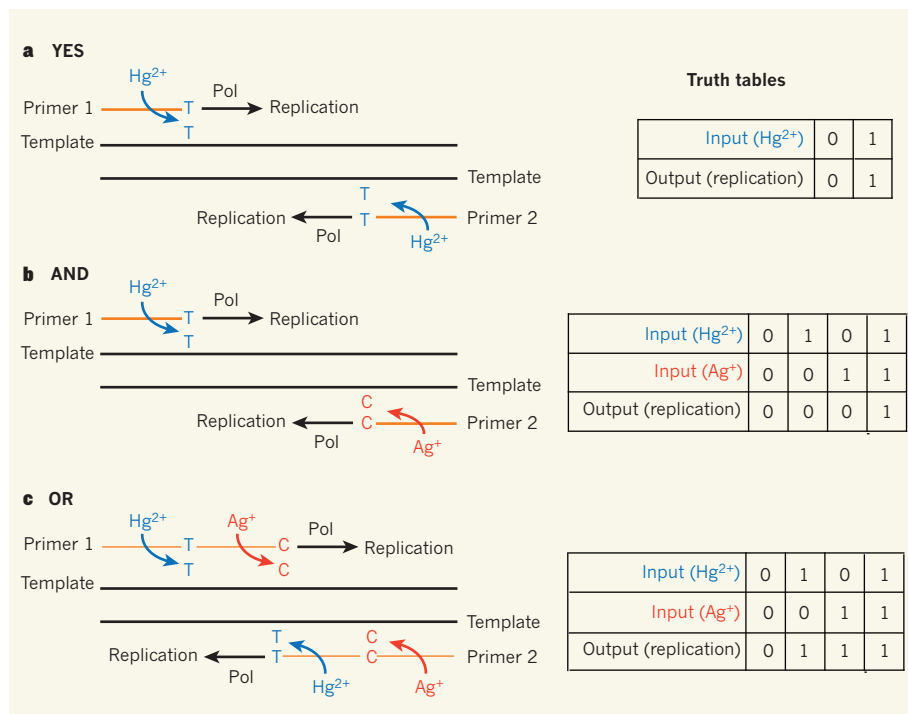
Figure 1 | DNA logic gates. In Park and colleagues' DNA-based logic gates[1], specially designed DNA primers generate a sequence mismatch (either thymine–thymine, T–T, or cytosine–cytosine, C–C) when they bind to an otherwise complementary DNA template. The mismatches prevent the primers from triggering DNA replication by a polymerase enzyme (Pol), but this is overcome in the presence of appropriate metal ions. The schemes depict conditions that lead to replication of each strand of the DNA template; both strands must be replicated for a positive output. The truth tables show all possible input combinations and resulting outputs for each gate. **a,** In the YES gate, both primers contain T–T mismatches and trigger DNA replication only in the presence of mercury ions (Hg^{2+}). **b,** The AND gate requires two different primers, one that generates a T–T mismatch and another that generates a C–C mismatch. DNA is replicated only when Hg^{2+} ions and silver ions (Ag^+, which override the C–C mismatch) are present. **c,** The OR gate requires both types of primer to be bound to each template strand. DNA replication occurs if at least one kind of metal ion is present.

authors also needed a way to override that block in response to a specific signal.

Metal ions provided the solution. It has been known for some time that mercury ions, Hg^{2+}, insert into T–T mismatches (where T represents the DNA base thymine), holding the bases tightly together[2,3]. Silver ions, Ag^+, bind to C–C mismatches in a similar way[4] (C is cytosine). Park et al.[1] found that the resulting T–Hg^{2+}–T and C–Ag^+–C interactions are highly specific and are not disturbed by the presence of other metal ions. Most importantly, these stable interactions can be thought of as artificial base-pairs that compensate for the original mismatch. Remarkably, the authors found that the presence of appropriate metal ions allowed their modified primers to initiate DNA replication, because the formation of metal 'bridges' across a C–C or T–T mismatch creates an artificial base-pair that is accepted by polymerase. They therefore used the binding of metal ions to these mismatches as robust, specific input signals for their DNA-based logic gates.

Using an ingenious combination of mismatch-containing primers and metal ions, Park et al. created the four Boolean logic gates

YES, PASS1, AND and OR. One of the simplest of these is YES, which produces a positive output in the presence of a positive input. The authors' DNA version of a YES gate involved two primers, each designed to generate a T–T mismatch with a different strand of the template (Fig. 1a). In this system, DNA replication (the output) occurred in the presence of Hg^{2+} ions (the input), but not in the ions' absence. The authors detected DNA replication using electrophoresis and fluorescence-based methods. PASS1 gates are also very simple — they provide a positive output whatever the input. The authors implemented a DNA PASS1 gate by simply changing the polymerase in their YES gate to an enzyme that replicates DNA even in the presence of mismatches, and irrespective of the presence of Hg^+ ions.

The authors' next target was more complex — an AND gate, which provides a positive output signal only if it receives two positive input signals. Their DNA system used two primers, one that generates a T–T mismatch with one strand of the template, and another that generates a C–C mismatch at the other strand of the template (Fig. 1b). As expected, the authors observed DNA replication

using this gate only in the presence of both Hg^{2+} and Ag^+ ions (which represent the two positive inputs).

Perhaps the most interesting challenge was the construction of an OR gate, which gives a positive output if at least one of the inputs is positive. For this, Park et al. had to design a system that replicates DNA in the presence of Ag^+ ions, or of Hg^{2+} ions, or both. They therefore constructed two pairs of primers, so that each pair bound sequentially to a different strand of the template DNA (Fig. 1c). Within each pair, one of the primers generated a T–T mismatch, whereas the other generated a C–C mismatch. Under the experimental conditions, DNA replication started from the T–T mismatch primers in the presence of Hg^{2+} ions, or from the C–C mismatched primers in the presence of Ag^+ ions. In the presence of both metal ions, either primer could trigger replication — as required for a functional OR gate.

Such DNA-based logic gates could be used to detect metal ions. But whereas a traditional sensor simply reports the presence of metal ions, logic gates could raise an alarm only if certain conditions are satisfied — reporting the presence of a specific ion only if a second ion is also present or absent, for example. Of course, smart sensors that rely on Ag^+ and Hg^{2+} ions alone as input signals would be very limiting. But the principles of Park and colleagues' design could form the basis of molecular logic gates that detect more complex signals, such as medicinally useful RNA fragments[5].

For applications in analytical devices that can judge complex situations, new artificial base pairs that can trigger polymerases need to be designed. Output signals also need to be devised that allow a more rapid readout of DNA replication (or of some other DNA-based positive output). Finally, systems of DNA logic gates must be established, in which the output of one gate is used as the input for another. This would provide a foundation for true DNA-based computers, perhaps heralding the dawn of molecular computing devices. Such devices are clearly a long way off, but the benefits would be substantial — it has been estimated that molecular computers would be orders of magnitude faster and more energy efficient than today's fastest computers[6,7]. ■

Thomas Carell is in the Department of Chemistry, Ludwig-Maximilians University, Butenandtstrasse 5-13, Munich 81377, Germany.
e-mail: thomas.carell@cup.uni-muenchen.de

1. Park, K. S., Jung, C. & Park, H. G. Angew. Chem. Int. Edn **49,** 9757–9760 (2010).
2. Katz, S. J. Am. Chem. Soc. **74,** 2238–2245 (1952).
3. Kuklenyik, Z. & Marzilli, L. G. Inorg. Chem. **35,** 5654–5662 (1996).
4. Ono, A. et al. Chem. Commun. 4825–4827 (2008).
5. Benenson, Y. et al. Nature **429,** 423–429 (2004).
6. Adleman, L. M. Science **266,** 1021–1024 (1994).
7. Chen, X. & Ellington, A. D. Curr. Opin. Biotechnol. **21,** 392–400 (2010).

ARTICLE

Reversing EphB2 depletion rescues cognitive functions in Alzheimer model

Moustapha Cissé[1,2], Brian Halabisky[1,2], Julie Harris[1,2], Nino Devidze[1], Dena B. Dubal[1,2], Binggui Sun[1,2], Anna Orr[1,2], Gregor Lotz[1,2], Daniel H. Kim[1], Patricia Hamto[1], Kaitlyn Ho[1], Gui-Qiu Yu[1] & Lennart Mucke[1,2]

Amyloid-β oligomers may cause cognitive deficits in Alzheimer's disease by impairing neuronal NMDA-type glutamate receptors, whose function is regulated by the receptor tyrosine kinase EphB2. Here we show that amyloid-β oligomers bind to the fibronectin repeats domain of EphB2 and trigger EphB2 degradation in the proteasome. To determine the pathogenic importance of EphB2 depletions in Alzheimer's disease and related models, we used lentiviral constructs to reduce or increase neuronal expression of EphB2 in memory centres of the mouse brain. In nontransgenic mice, knockdown of EphB2 mediated by short hairpin RNA reduced NMDA receptor currents and impaired long-term potentiation in the dentate gyrus, which are important for memory formation. Increasing EphB2 expression in the dentate gyrus of human amyloid precursor protein transgenic mice reversed deficits in NMDA receptor-dependent long-term potentiation and memory impairments. Thus, depletion of EphB2 is critical in amyloid-β-induced neuronal dysfunction. Increasing EphB2 levels or function could be beneficial in Alzheimer's disease.

Soluble amyloid-β oligomers may contribute to learning and memory deficits in Alzheimer's disease by inhibiting NMDA-receptor-dependent long-term potentiation (LTP)[1–3], thought to underlie memory formation[4]. In Alzheimer's disease, hippocampal NMDA-receptor-subunit levels are reduced[5], and protein levels and the phosphorylation status of NMDA-receptor subunits NR1, NR2A and NR2B correlate with cognitive performance[6]. Human amyloid precursor protein (hAPP) transgenic mice with high brain levels of amyloid-β oligomers have reduced hippocampal levels of tyrosine-phosphorylated NMDA receptors and key components of NMDA-receptor-dependent signalling pathways[7,8]. Alzheimer's disease patients and hAPP mice have hippocampal depletions of the receptor tyrosine kinase EphB2[9], which regulates NMDA-receptor trafficking and function by interacting with NMDA receptors and Src-mediated tyrosine phosphorylation[10–13]. EphB2 regulates NMDA-receptor-dependent Ca^{2+} influx and downstream transcription factors involved in LTP formation[12], such as Fos, which is depleted in the dentate gyrus of hAPP mice. Mice lacking EphB2[10,14] or Fos[15] have impaired NMDA-receptor-dependent LTP and memory deficits. We hypothesized that EphB2 depletion in Alzheimer's disease-related models is caused by amyloid-β oligomers and that reductions in EphB2 contribute to amyloid-β-induced deficits in synaptic plasticity and cognitive functions (Supplementary Fig. 1). Here we confirm these hypotheses and show that reversing EphB2 depletion in the dentate gyrus of hAPP mice reverses LTP and memory impairments.

Amyloid-β oligomers bind to EphB2

To determine if amyloid-β oligomers interact directly with EphB2, we measured binding of biotinylated synthetic amyloid-β1–42 oligomers to a purified recombinant EphB2–Fc chimaera. Biotinylated amyloid-β oligomers and EphB2–Fc were pulled down together by avidin agarose beads (Supplementary Fig. 2a, b) and co-immunoprecipitated under cell-free conditions (Supplementary Fig. 2c, d). EphB2 and amyloid-β oligomers also co-immunoprecipitated from homogenates of primary neurons (Supplementary Fig. 2e–g). Thus, amyloid-β oligomers may interact directly with the extracellular region of EphB2.

This region comprises a ligand-binding (LB) domain, a cysteine-rich (CR) domain, and a fibronectin type III repeats (FN) domain (Fig. 1a). To determine which domain mediates the interaction with amyloid-β oligomers, we generated EphB2–GST deletion mutants lacking the LB domain (ΔLB-EphB2) or the FN domain (ΔFN-EphB2) (Fig. 1a). Amyloid-β oligomers bound to FL-EphB2 and ΔLB-EphB2 but not ΔFN-EphB2 (Fig. 1b, c), indicating that the FN domain is critical for their interaction with EphB2.

Deleting the FN domain did not affect trafficking of EphB2 to the cell surface (Supplementary Fig. 3a). FL-EphB2 and ΔFN-EphB2 both phosphorylated the NMDA-receptor subunit NR1 after stimulation with the EphB2 ligand, Fc-ephrin-B2 (Supplementary Fig. 3b–d). Thus, deleting the FN domain did not eliminate the kinase function of EphB2. Deleting the LB domain prevented Fc-ephrin-B2-induced phosphorylation of NR1 (Supplementary Fig. 3b–d).

Mechanisms of amyloid-β-induced EphB2 depletion

At 3–4 but not 2 months of age, EphB2 messenger RNA and protein levels in the hippocampus were lower in hAPP mice than in nontransgenic controls, and were lower in humans with Alzheimer's disease than in nondemented controls (data not shown), consistent with previous findings[9].

As reported by others[16], we observed a doublet of putative EphB2 carboxy-terminal fragments (CTFs) of 45–50 kDa in hippocampi of hAPP mice and nontransgenic controls on western blots (not shown). Relative to nontransgenic controls, hAPP mice showed a comparable decrease in CTFs and FL-EphB2 (not shown) and no difference in the ratio of CTF1+CTF2:FL-EphB2 (hAPP, 2.7 ± 0.36; nontransgenic, 2.3 ± 0.59; $P = 0.55$ by t test). Thus, pathologically raised levels of amyloid-β do not affect EphB2 cleavage into CTFs.

Treating primary neuronal cultures from wild-type rats with naturally secreted amyloid-β oligomers caused severe EphB2 depletions by 3 days (Fig. 1d–f). Amyloid-β oligomers reduced EphB2 mRNA levels (Fig. 1g), but the reduction was subtle and unlikely to account for the severe EphB2 protein depletion.

Amyloid-β-induced depletion of EphB2 was blocked by the proteasome inhibitor lactacystin (Fig. 1h, i). Bafilomycin, an inhibitor of

[1]Gladstone Institute of Neurological Disease, San Francisco, California 94158, USA. [2]Department of Neurology, University of California, San Francisco, California 94158, USA.
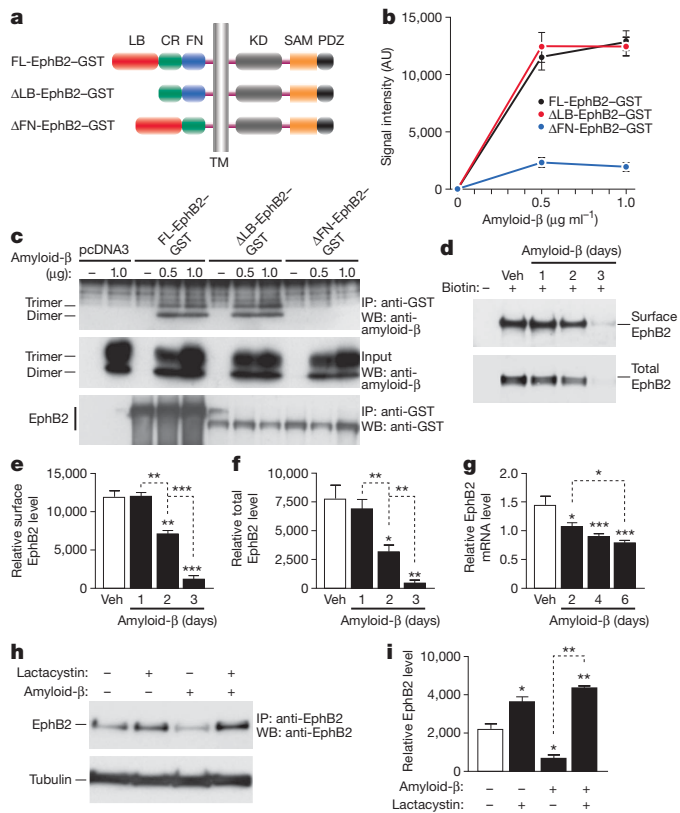
Figure 1 | Amyloid-β oligomers bind to the fibronectin repeats domain of EphB2 and cause degradation of EphB2 in the proteasome. a, Domain structure of full-length (FL) EphB2 and deletion constructs. Ligand-binding (LB) domain, cysteine-rich (CR) region, fibronectin type III repeats (FN) domain, transmembrane (TM) region, tyrosine kinase (KD) domain, sterile alpha motif (SAM) domain, and PSD95, DLG and ZO1 (PDZ) domain. **b,** Binding of amyloid-β dimers and trimers to different EphB2 constructs. See Supplementary Table 2 for experimental details pertaining to data shown in figures. AU, arbitrary units. **c,** Representative western blot (WB). IP, immunoprecipitation. **d–f,** Amyloid-β-induced depletion of EphB2. Primary rat neurons were treated with amyloid-β or vehicle (Veh), and surface and total levels of EphB2 were determined by western blots. Representative western blots are shown in **d. e, f,** Quantification of surface (**e**) and total (**f**) levels of EphB2. **g,** EphB2 mRNA levels in primary neurons treated with amyloid-β or vehicle. **h, i,** Lactacystin blocks amyloid-β-induced depletion of EphB2 in primary neurons. Representative western blot (**h**) and quantification of signals (**i**). For all experiments, $n = 3–6$ wells per condition from three independent experiments. *$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$ versus empty bars or as indicated by brackets (Tukey test). Values are means ± s.e.m.

endosomal acidification, had no effect (Supplementary Fig. 4b, c). Compared with amyloid-β treatment alone, treatment of cells with lactacystin alone or together with amyloid-β increased ubiquitinated EphB2 (Supplementary Fig. 4a). These results indicate that amyloid-β depletes neuronal EphB2 mainly by enhancing its proteasomal degradation.

EphB2 depletion impairs synaptic plasticity

To determine if EphB2 depletion interferes with NMDA-receptor-dependent functions, we generated lentiviral vectors expressing green fluorescent protein (GFP) and anti-EphB2 shRNA (Lenti-sh-EphB2–GFP) or scrambled control shRNA (Lenti-sh-SCR–GFP). In neuronal cultures, Lenti-sh-EphB2–GFP reduced EphB2 mRNA and protein levels (Fig. 2a, b) and surface levels of NR1 (Fig. 2c–e). In cultures co-infected with a mutant EphB2 construct whose mRNA is resistant to sh-EphB2 (Lenti-mut-EphB2–Flag) and Lenti-sh-EphB2–GFP, EphB2 and surface NR1 were not reduced (Supplementary Fig. 5), excluding an off-target effect. Next we examined the effects of sh-EphB2 on expression of the immediate-early gene c-fos, which depends on NMDA receptors and is regulated by EphB2[12]. Anti-EphB2 shRNA prevented Fc-ephrin-B2-induced increases in Fos expression in
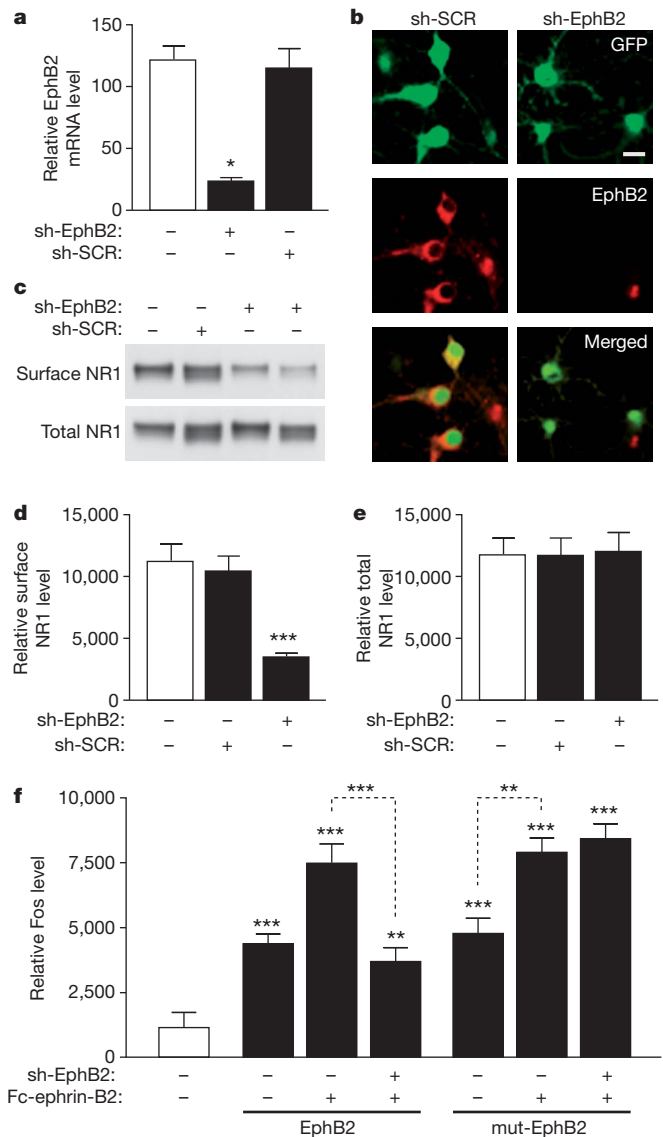
Figure 2 | Knockdown of EphB2 reduces surface NR1 levels and Fc-ephrin-B2-dependent Fos expression. a, b, EphB2 expression is reduced in primary neurons infected with Lenti-sh-EphB2–GFP as determined by RT–qPCR (**a**) or EphB2 immunostaining (**b**). Scale bar, 20 μm. **c–e,** Reduction of EphB2 levels by Lenti-sh-EphB2–GFP and effect on surface NR1 levels. **f,** shRNA against wild-type but not mutated EphB2 reduces Fc-ephrin-B2-dependent Fos expression. Primary rat neurons were co-infected or not with Lenti-sh-EphB2–GFP (sh-EphB2) in combination with either Lenti-EphB2 encoding wild-type EphB2 or Lenti-mut-EphB2 (mut-EphB2) encoding a mutated EphB2 mRNA that is not recognized by sh-EphB2. Four days later, cells were stimulated with clustered multimeric recombinant Fc-ephrin-B2 ligand to activate EphB2. $n = 3–6$ wells per condition from three independent experiments. *$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$ versus empty bar or as indicated by brackets (Tukey's test). Values are means ± s.e.m.

neurons expressing wild-type EphB2, but not in neurons expressing mutant EphB2 (Fig. 2f). Thus, depleting EphB2 reduces surface NR1 expression and impairs NMDA-receptor-dependent gene expression.

To explore whether EphB2 depletion accounts for LTP deficits in hAPP mice[8], we reduced EphB2 in the dentate gyrus of nontransgenic mice. Although granule cells are not very susceptible to degeneration in Alzheimer's disease, perforant path to granule cell synapses are affected early and severely[17,18].

Two anti-EphB2 shRNAs reduced EphB2 mRNA and protein levels in neuronal culture (Supplementary Fig. 6). Mice injected with lentiviral vectors expressing sh-EphB2-308–GFP (Fig. 3a, b) or sh-EphB2-306–GFP (Supplementary Fig. 7a, b) had lower EphB2 mRNA levels in the

dentate gyrus than controls. Transduction efficiencies (Supplementary Fig. 8) were 50–74% (mean ± s.e.m., 62.4 ± 6.2; $n = 7$ mice), consistent with other reports[19,20].

Field (Fig. 3c) and whole-cell patch-clamp recordings (Fig. 3e) from dentate gyrus granule cells in acute hippocampal slices from Lenti-sh-EphB2–GFP-injected nontransgenic mice revealed prominent LTP deficits similar to those in untreated hAPP J20 (Fig. 3d, f) and other lines of hAPP mice[21,22]. Lenti-sh-SCR–GFP-injected nontransgenic mice had robust LTP in the dentate gyrus (Fig. 3c, e). Whole-cell recordings from individual GFP-negative granule cells in Lenti-sh-ephB2–GFP-injected mice revealed no LTP deficits, compared with GFP-negative granule cells in untreated nontransgenic mice and GFP-positive granule cells in Lenti-sh-SCR–GFP-injected mice ($P > 0.1$ by repeated-measures ANOVA, $n = 6$ neurons from 3 mice per group; data not shown).

EphB2 depletion reduces synaptic strength

LTP at the medial perforant path to granule cell synapses depends on NMDA-receptor activity[23]. We determined whether impaired synaptic

plasticity in sh-EphB2-treated nontransgenic and untreated hAPP mice is related to a selective impairment of these glutamate receptors. NMDA-receptor-mediated, but not α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-receptor-mediated synaptic transmission strength at this synapse was affected in sh-EphB2-treated nontransgenic mice (Fig. 3g) and untreated hAPP mice (Fig. 3h), as determined by field recordings and analysis of input-output (I/O) curves. These alterations markedly reduced ratios of NMDA-receptor-to AMPA-receptor-mediated synaptic strength in both groups (Fig. 3j). Similar results were obtained by whole-cell recordings from individual granule cells (Fig. 3i, k). To exclude a contribution of alterations in AMPA-receptor currents to the altered ratios, we recorded pharmacologically isolated, AMPA-receptor-mediated miniature excitatory synaptic currents (mEPSCs). The four groups of mice had comparable mEPSC peak amplitudes (Supplementary Fig. 9). Thus, like amyloid-β, EphB2 depletion probably reduces LTP by impairing NMDA-receptor function.

EphB2 rescues synaptic functions in hAPP mice

To determine if increasing EphB2 expression in the dentate gyrus of hAPP mice reverses LTP deficits, we used a lentivirus expressing EphB2–Flag (Lenti-EphB2–Flag). Lenti-EphB2–Flag-treated hAPP and nontransgenic mice had comparable EphB2–Flag expression levels in the dentate gyrus (Fig. 4a and Supplementary Fig. 10). Lenti-empty-treated nontransgenic mice and Lenti-EphB2–Flag-treated hAPP mice had comparable dentate gyrus levels of total (endogenous and exogenous) EphB2 (Fig. 4b), indicating that EphB2 levels in hAPP mice were normalized. EphB2 levels were lower in Lenti-empty-injected hAPP mice and higher in Lenti-EphB2–Flag-injected nontransgenic mice (Fig. 4b). Increasing dentate gyrus EphB2 levels reversed LTP deficits in two independent cohorts of hAPP mice but did not alter LTP in nontransgenic mice (Fig. 4c).

Lenti-EphB2–Flag-treated mice showed a trend towards lower amyloid-β levels in the dentate gyrus (Supplementary Fig. 11), but this trend did not reach statistical significance. At analysis, hAPP mice were 4–5-months old and had not yet formed plaques, excluding EphB2 effects on plaque formation. To determine if LTP rescue was due to improved NMDA-receptor function, we again measured AMPA-receptor- and NMDA-receptor-mediated synaptic strength. Increasing EphB2 levels in the dentate gyrus of hAPP mice reversed deficits in NMDA-receptor-mediated synaptic strength without
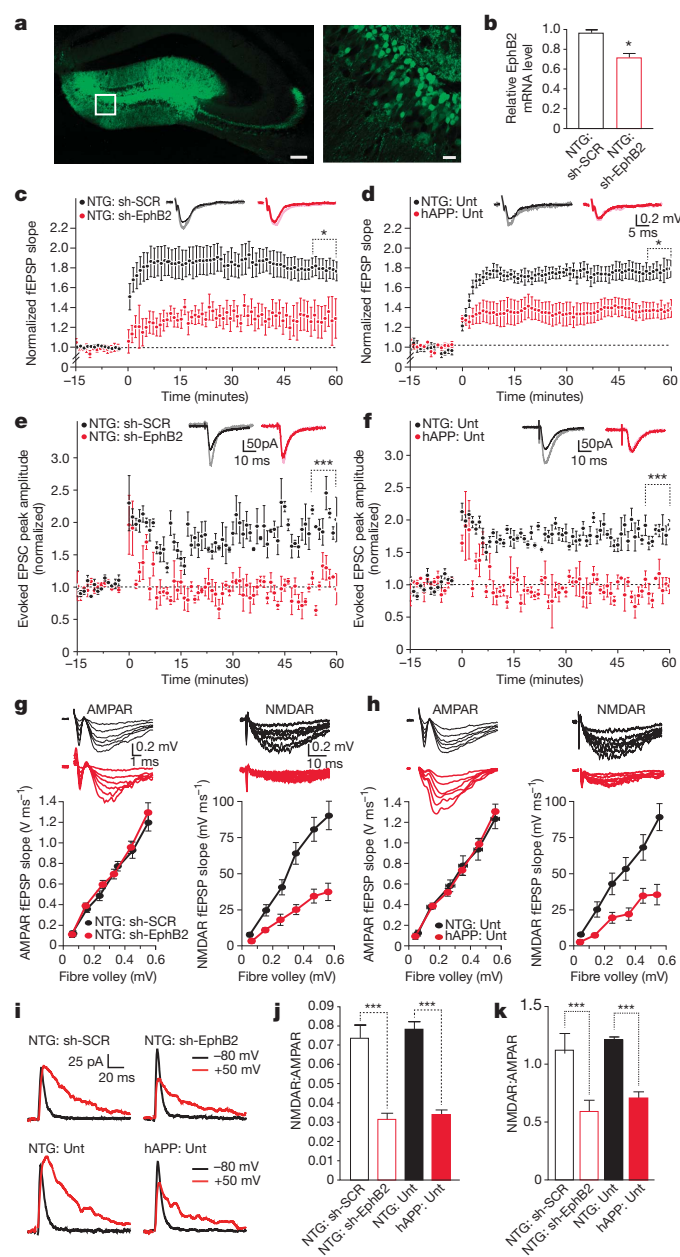
Figure 3 | Knockdown of EphB2 reduces LTP in dentate gyrus granule cells of nontransgenic mice. a, Anti-GFP immunostaining of dentate gyrus showing infected neurons in Lenti-sh-EphB2–GFP injected mice. Right panel shows higher magnification image of boxed region on left. Scale bars: 100 μm (left), 25 μm (right). **b,** EphB2 mRNA levels in the entire dentate gyrus (reflecting levels in infected and uninfected cells) ($n = 5$–7 mice per condition). *$P < 0.001$ versus sh-SCR (t test). NTG, nontransgenic mice. **c–f,** LTP at the medial perforant path to granule cell synapse measured by field recordings (**c, d**) or by whole-cell patch clamp from individual GFP-positive cells (**e, f**) in the dentate gyrus. LTP was impaired in nontransgenic mice treated with Lenti-sh-EphB2–GFP (sh-EphB2) compared to nontransgenic mice treated with Lenti-sh-SCR–GFP (sh-SCR) (**c, e**). Similar LTP impairments were observed in untreated (Unt) hAPP mice (**d, f**) (NTG: sh-EphB2 versus hAPP: Unt). *$P < 0.05$, ***$P < 0.001$ (repeated-measures ANOVA and Bonferroni post-hoc test on the last 10 min of data). $n = 8$–9 slices from 3–4 mice per treatment (**c**) or genotype (**d**). **g, h,** Comparison of AMPA-receptor (AMPAR)-mediated (left) and NMDA-receptor (NMDAR)-mediated (right) input-output (I/O) relationships in the medial perforant path to granule cell synapses of nontransgenic mice treated with sh-EphB2 versus sh-SCR (**g**) and of untreated nontransgenic (NTG: Unt) versus hAPP (hAPP: Unt) mice (**h**). **i,** Example traces of evoked glutamate receptor currents from individual granule cells voltage clamped at –80 or 50 mV to measure AMPA-receptor- and NMDA-receptor-mediated currents, respectively. **j, k,** Summary plot of the ratios of NMDA-receptor I/O relationships to AMPA-receptor I/O relationships measured by field recordings (**j**) or by individual granule cells (**k**). ***$P < 0.001$ (two-way ANOVA and Bonferroni post-hoc test). $n = 8$–9 slices from 3–4 mice per group. Values are means ± s.e.m.
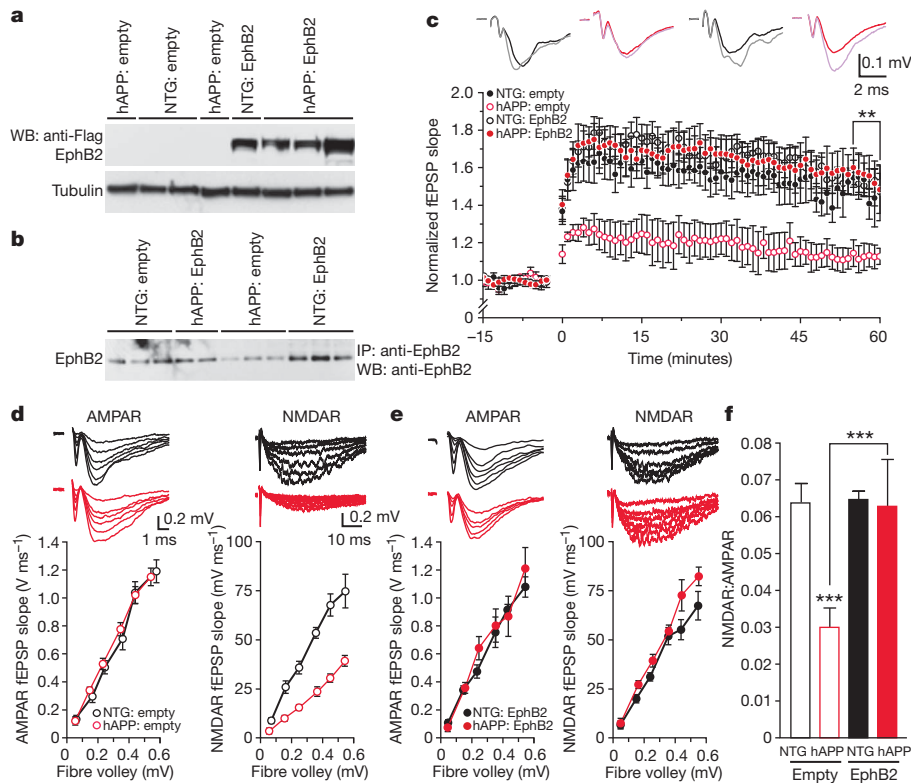
changing AMPA-receptor-mediated synaptic strength (Fig. 4d, e), normalizing the balance between them (Fig. 4f). Overexpressing EphB2 did not alter NMDA-receptor- or AMPA-receptor-mediated synaptic strength in nontransgenic mice (Fig. 4d–f).

Increasing EphB2 expression in granule cells did not reverse impairments in paired pulse modification at perforant path to granule cell synapses (Supplementary Fig. 12a) or in synaptic strength at Schaffer collateral to CA1 pyramidal cell synapses (Supplementary Fig. 12b, c).

EphB2 ameliorates cognitive deficits in hAPP mice

To determine if increasing EphB2 levels in the dentate gyrus also reverses learning and memory deficits in hAPP mice[24–27], we injected Lenti-EphB2–Flag or Lenti-empty bilaterally into the dentate gyrus of hAPP and nontransgenic mice and analysed them behaviourally 2 months later.

Spatial learning and memory in the Morris water maze is strongly affected by dentate gyrus impairments[28]. In the spatial, hidden-platform component, Lenti-EphB2–Flag-treated but not Lenti-empty-treated hAPP mice performed at control levels (Fig. 5a, b). Overexpressing EphB2 did not alter learning in nontransgenic mice (Fig. 5a, b). All groups of mice learned similarly well in the cued-platform component (data not shown).

In a probe trial, Lenti-empty-treated but not Lenti-EphB2–Flag-treated hAPP mice took longer to reach the original platform location than Lenti-empty-treated nontransgenic controls (Fig. 5c). Lenti-EphB2–Flag-treated nontransgenic mice performed slightly worse than Lenti-empty-treated nontransgenic mice (Fig. 5c) ($P = 1.0$ by one-way ANOVA and Bonferroni post-hoc test).

In the novel object recognition test, Lenti-EphB2-treated but not Lenti-empty-treated hAPP mice spent more time exploring the novel object (Fig. 5d). In the novel place recognition test, Lenti-EphB2-treated but not Lenti-empty-treated hAPP mice spent more time exploring the object whose location had changed (Fig. 5e). Thus, increasing EphB2 expression in the dentate gyrus of hAPP mice ameliorates deficits in both spatial and nonspatial learning and memory.

Finally, we assessed passive avoidance learning, which depends at least partly on hippocampal functions[29,30]. During training, escape

latencies were similar across groups (Fig. 5f). However, 24 h later, Lenti-empty-treated hAPP mice were severely impaired, whereas all other groups performed well (Fig. 5f). Increasing dentate gyrus EphB2 levels in hAPP mice did not reverse behavioural deficits that were probably caused by impairments of other brain regions, including hyperactivity in the open field and disinhibition in the elevated plus maze (Supplementary Fig. 13).

Discussion

Our study shows that EphB2 depletion contributes to amyloid-β-induced neuronal deficits and cognitive dysfunction. Reducing neuronal EphB2 levels caused functional deficits similar to those caused by amyloid-β, including deficits in NMDA-receptor-dependent synaptic strength and gene expression and impaired LTP and memory. Increasing neuronal EphB2 levels in hAPP mice reversed these deficits, indicating that EphB2 impairment is necessary and sufficient to elicit them and that increasing EphB2 activity counteracts amyloid-β-induced neuronal dysfunction. Consistent with a previous report[9], EphB2 depletion in memory-related brain regions was detected not only in hAPP mice, but also in humans with Alzheimer's disease, underlining the potential clinical relevance of our findings. Our data further indicate that the depletion of EphB2 by amyloid-β oligomers involves direct binding of amyloid-β oligomers to the FN repeats domain of EphB2 and EphB2 degradation in the proteasome. Reduction of EphB2 mRNA may have an additional role.

Our results and those of others indicate that neuronal EphB2 depletion causes deficits in learning and memory by impairing NMDA-receptor functions (Supplementary Fig. 1). EphB2 modulates NMDA receptors by tyrosine phosphorylation and recruits active NMDA receptors to excitatory synapses[10–12]. EphB2-deficient mice have LTP deficits[10,14] and fewer NR1 subunits in the postsynaptic density[10]. Our results are consistent with these findings, although LTP deficits after shRNA knockdown of EphB2 in adult nontransgenic mice were more severe than those in EphB2-deficient mice. Other members of the large Eph family might partially compensate for EphB2 ablation during early development. The more severe deficits after acute EphB2 knockdown in adults probably reflect the lack of such compensation.
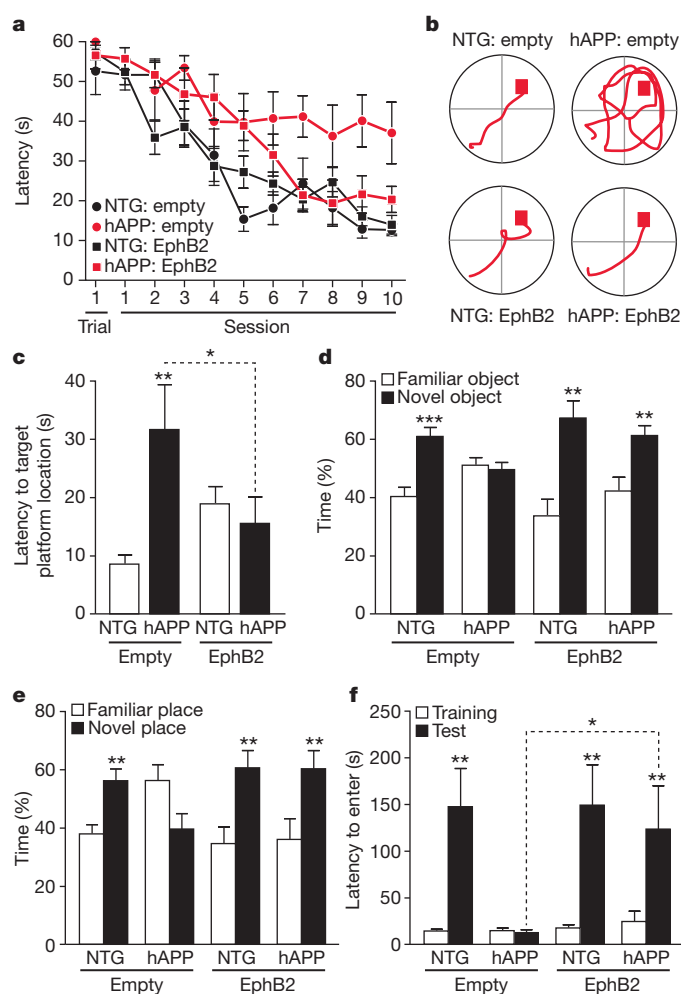
Figure 5 | Increasing EphB2 expression in the dentate gyrus ameliorates learning and memory deficits in hAPP mice. a, Learning curves during spatial training in the Morris water maze. The latency for each mouse to reach the hidden platform was recorded. Trial 1 represents performance on the first trial, and subsequent sessions represent the average of two training trials. Lenti-empty treated hAPP mice had longer latencies and travelled farther (not shown) to find the hidden platform than all other groups ($P < 0.0001$, repeated-measures ANOVA). **b**, Representative paths from the last session of hidden-platform training. **c**, Time it took mice to reach the target platform location during a probe trial (platform removed) 24 h after the last hidden-platform training. $*P < 0.05$, $**P < 0.01$ versus first bar or as indicated by bracket (one-way ANOVA followed by Bonferroni post-hoc test). **d**, Object recognition memory as reflected by the percentage of time mice spent exploring a familiar versus a novel object during a 10-min test session. $**P < 0.01$, $***P < 0.001$ versus familiar object (paired t test). **e**, Spatial location memory as reflected by the percentage of time mice spent exploring familiar objects whose locations were or were not altered. $**P < 0.01$ versus familiar place (t test). **f**, Passive avoidance memory. $*P < 0.05$, $**P < 0.01$, versus training or as indicated by bracket (one-way nonparametric Kruskal–Wallis test followed by Dunn's post test). $n = 9$–12 mice per genotype and treatment. Values are means ± s.e.m.

Modulation of other LTP-related proteins also results in different outcomes, depending on when it is initiated[31].

Amyloid-β may impair LTP by inducing internalization of NMDA receptors[32,33]. We found that depletion of EphB2 contributes to the amyloid-β-induced decrease in NMDA receptors and that increasing EphB2 expression markedly improves LTP and memory even in the presence of high amyloid-β levels. Increased EphB2 levels probably increase surface NMDA-receptor expression. Indeed, increasing neuronal EphB2 expression reversed amyloid-β-induced deficits in NMDA-receptor-mediated synaptic strength.

Opposition of amyloid-β-induced surface depletion of NMDA receptors is the most parsimonious interpretation of the EphB2-mediated

rescue effects (Supplementary Fig. 1). However, amyloid-β may impair LTP and memory through alternative processes, and increased expression of EphB2 may counteract amyloid-β effects also through downstream signalling mechanisms.

Manipulating individual functional hubs of neurons can profoundly affect a larger network[34,35]. Even manipulating an individual neuron can affect the global brain state[36]. Thus, improving the function of a subset of neurons might allow an impaired brain region to better support specific behaviours. The current study supports this hypothesis: increasing EphB2 expression in a subset of granule cells improved dentate gyrus LTP and learning and memory in hAPP mice. It remains to be determined whether EphB2 depletions contribute to amyloid-β-dependent impairments in other brain regions and whether increasing neuronal EphB2 levels in these regions is tolerated as well as it was in the dentate gyrus. If so, pharmacological treatments might be used to increase EphB2 expression or activity. Our results indicate additional entry points for interventions (Supplementary Fig. 1). For example, it may be possible to identify small molecules that block the binding of amyloid-β oligomers to EphB2's FN repeats domain, prevent proteasomal degradation of EphB2, or improve its interactions with NMDA receptors.

METHODS SUMMARY

General. Unless indicated otherwise, all data reported in this paper were obtained in blind-coded experiments, in which the investigators who obtained the data were unaware of the specific genotype and treatment of mice, brain slices and cell cultures. For number of mice, slices and cell cultures analysed in each experiment, refer to Supplementary Table 1. For experimental details related to each figure legend, refer to Supplementary Table 2.

Experimental models. Heterozygous transgenic and nontransgenic mice were from hAPP line J20[7,8,37,38]. Primary neuronal cultures from wild-type rats were treated with medium conditioned by CHO cells that do or do not produce human amyloid-β oligomers (Supplementary Figs 14 and 15 and refs 39, 40.).

Experimental manipulations. Lentiviral constructs directing neuronal expression of no transgene products, EphB2–Flag, or GFP in combination with anti-EphB2 shRNAs or scrambled control shRNA were injected stereotactically into the dentate gyrus of mice[20,41]. Neuronal cultures were infected with some of these constructs and stimulated with Fc-ephrin-B2 or Fc control[12,42].

Outcome measures. The interaction between biotinylated or naturally secreted amyloid-β oligomers and EphB2 was assessed under cell-free conditions and in neuronal cultures of primary neurons or HEK cells by pull-down with avidin agarose beads[43] or immunoprecipitation and western blot[44]. EphB2 and NR1 levels in brain tissues or neuronal cultures were determined by immunoprecipitation and western blot or western blot alone[44]. Corresponding transcripts were measured by quantitative polymerase chain reaction with reverse transcription (RT–qPCR). Fos expression in neuronal cultures was determined by western blot[44]. Field recordings[8] or whole-cell patch-clamp recordings[45] from acute hippocampal slices were used to determine synaptic strength (fEPSP I/O relationships; mediated by either AMPA receptors or NMDA receptors), synaptic plasticity (LTP), and NMDA-receptor:AMPA-receptor ratios of EPSCs at the medial perforant path to dentate gyrus granule cell synapses. Learning and memory were assessed in the Morris water maze, novel object recognition test, novel place recognition test, and passive avoidance test[46–49]. Amyloid-β levels in the dentate gyrus of hAPP-J20 mice were determined by ELISA[50].

1. Walsh, D. M. & Selkoe, D. J. Deciphering the molecular basis of memory failure in Alzheimer's disease. *Neuron* **44,** 181–193 (2004).
2. Shankar, G. M. *et al.* Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature Med.* **14,** 837–842 (2008).
3. Kamenetz, F. *et al.* APP processing and synaptic function. *Neuron* **37,** 925–937 (2003).
4. Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44,** 5–21 (2004).
5. Ikonomovic, M. D. *et al.* Distribution of glutamate receptor subunit NMDAR1 in the hippocampus of normal elderly and patients with Alzheimer's disease. *Exp. Neurol.* **160,** 194–204 (1999).
6. Sze, C., Bi, H., Kleinschmidt-DeMasters, B. K., Filley, C. M. & Martin, L. J. *N*-Methyl-D-aspartate receptor subunit proteins and their phosphorylation status are altered selectively in Alzheimer's disease. *J. Neurol. Sci.* **182,** 151–159 (2001).

7. Palop, J. J. *et al.* Vulnerability of dentate granule cells to disruption of Arc expression in human amyloid precursor protein transgenic mice. *J. Neurosci.* **25,** 9686–9693 (2005).

8. Palop, J. J. *et al.* Aberrant excitatory neuronal activity and compensatory remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron* **55,** 697–711 (2007).

9. Simon, A. M. *et al.* Early changes in hippocampal Eph receptors precede the onset of memory decline in mouse models of Alzheimer's disease. *J. Alzheimers Dis.* **17,** 773–786 (2009).

10. Henderson, J. T. *et al.* The receptor tyrosine kinase EphB2 regulates NMDA-dependent synaptic function. *Neuron* **32,** 1041–1056 (2001).

11. Dalva, M. B. *et al.* EphB receptors interact with NMDA receptors and regulate excitatory synapse formation. *Cell* **103,** 945–956 (2000).

12. Takasu, M. A., Dalva, M. B., Zigmond, R. E. & Greenberg, M. E. Modulation of NMDA receptor-dependent calcium influx and gene expression through EphB receptors. *Science* **295,** 491–495 (2002).

13. Chen, Y., Fu, A. K. & Ip, N. Y. Bidirectional signaling of ErbB and Eph receptors at synapses. *Neuron Glia Biol.* **4,** 211–221 (2008).

14. Grunwald, I. C. *et al.* Kinase-independent requirement of EphB2 receptors in hippocampal synaptic plasticity. *Neuron* **32,** 1027–1040 (2001).

15. Fleischmann, A. *et al.* Impaired long-term memory and NR2A-type NMDA receptor-dependent synaptic plasticity in mice lacking c-Fos in the CNS. *J. Neurosci.* **23,** 9116–9122 (2003).

16. Litterst, C. *et al.* Ligand binding and calcium influx induce distinct ectodomain/γ-secretase-processing pathways of EphB2 receptor. *J. Biol. Chem.* **282,** 16155–16163 (2007).

17. Wakabayashi, K., Honer, W. G. & Masliah, E. Synapse alterations in the hippocampal-entorhinal formation in Alzheimer's disease with and without Lewy body disease. *Brain Res.* **667,** 24–32 (1994).

18. Scheff, S. W. & Price, D. A. Alzheimer's disease-related alterations in synaptic density: neocortex and hippocampus. *J. Alzheimers Dis.* **9,** 101–115 (2006).

19. Mueller-Steiner, S. *et al.* Anti-amyloidogenic and neuroprotective functions of cathepsin B: implications for Alzheimer's disease. *Neuron* **51,** 703–714 (2006).

20. Sun, B. *et al.* Imbalance between GABAergic and glutamatergic transmissions impairs adult neurogenesis in an animal model of Alzheimer's disease. *Cell Stem Cell* **5,** 624–633 (2009).

21. Shemer, I. *et al.* Non-fibrillar β-amyloid abates spike-timing-dependent synaptic potentiation at excitatory synapses in layer 2/3 of the neocortex by targeting postsynaptic AMPA receptors. *Eur. J. Neurosci.* **23,** 2035–2047 (2006).

22. Ashe, K. H. & Zahs, K. R. Probing the biology of Alzheimer's disease in mice. *Neuron* **66,** 631–645 (2010).

23. Colino, A. & Malenka, R. C. Mechanisms underlying induction of long-term potentiation in rat medial and lateral perforant paths *in vitro. J. Neurophysiol.* **69,** 1150–1159 (1993).

24. Harris, J. A. *et al.* Many neuronal and behavioral impairments in transgenic mouse models of Alzheimer's disease are independent of caspase cleavage of the amyloid precursor protein. *J. Neurosci.* **30,** 372–381 (2010).

25. Sanchez-Mejia, R. O. *et al.* Phospholipase A2 reduction ameliorates cognitve deficits in mouse model of Alzheimer's disease. *Nature Neurosci.* **11,** 1311–1318 (2008).

26. Meilandt, W. J. *et al.* Enkephalin elevations contribute to neuronal and behavioral impairments in a transgenic mouse model of Alzheimer's disease. *J. Neurosci.* **28,** 5007–5017 (2008).

27. Roberson, E. D. *et al.* Reducing endogenous tau ameliorates amyloid β-induced deficits in an Alzheimer's disease mouse model. *Science* **316,** 750–754 (2007).

28. Nguyen, P. V., Abel, T., Kandel, E. R. & Bourtchouladze, R. Strain-dependent differences in LTP and hippocampus-dependent memory in inbred mice. *Learn. Mem.* **7,** 170–179 (2000).

29. Nakajima, R. *et al.* Comprehensive behavioral phenotyping of calpastatin-knockout mice. *Mol. Brain* **1,** 7 (2008).

30. Potter, M. C. *et al.* Reduction of endogenous kynurenic acid formation enhances extracellular glutamate, hippocampal plasticity, and cognitive behavior. *Neuropsychopharmacology* **35,** 1734–1742 (2010).

31. Terashima, A. *et al.* An essential role for PICK1 in NMDA receptor-dependent bidirectional synaptic plasticity. *Neuron* **57,** 872–882 (2008).

32. Snyder, E. M. *et al.* Regulation of NMDA receptor trafficking by amyloid-β. *Nature Neurosci.* **8,** 1051–1058 (2005).

33. Kurup, P. *et al.* Aβ-mediated NMDA receptor endocytosis in Alzheimer's disease involves ubiquitination of the tyrosine phosphatase STEP61. *J. Neurosci.* **30,** 5948–5957 (2010).

34. Bonifazi, P. *et al.* GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science* **326,** 1419–1424 (2009).

35. Han, J. H. *et al.* Selective erasure of a fear memory. *Science* **323,** 1492–1496 (2009).

36. Li, C. Y., Poo, M. M. & Dan, Y. Burst spiking of a single cortical neuron modifies global brain state. *Science* **324,** 643–646 (2009).

37. Rockenstein, E. M. *et al.* Levels and alternative splicing of amyloid β protein precursor (APP) transcripts in brains of transgenic mice and humans with Alzheimer's disease. *J. Biol. Chem.* **270,** 28257–28267 (1995).

38. Mucke, L. *et al.* High-level neuronal expression of Aβ$_{1-42}$ in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20,** 4050–4058 (2000).

39. Koo, E. H. & Squazzo, S. L. Evidence that production and release of amyloid β-protein involves the endocytic pathway. *J. Biol. Chem.* **269,** 17386–17389 (1994).

40. Walsh, D. M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo. Nature* **416,** 535–539 (2002).

41. Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates.* (Academic, 1997).

42. Xia, Z., Dudek, H., Miranti, C. K. & Greenberg, M. E. Calcium influx via the NMDA receptor induces immediate early gene transcription by a MAP kinase/ERK-dependent mechanism. *J. Neurosci.* **16,** 5425–5436 (1996).

43. Laurén, J., Gimbel, D. A., Nygaard, H. B., Gilbert, J. W. & Strittmatter, S. M. Cellular prion protein mediates impairment of synaptic plasticity by amyloid-β oligomers. *Nature* **457,** 1128–1132 (2009).

44. Alfa Cisse, M. *et al.* M1 and M3 muscarinic receptors control physiological processing of cellular prion by modulating Alzheimer's disease AM17 phosphorylation and activity. *J. Neurosci.* **27,** 4083–4092 (2007).

45. Wu, J., Rush, A., Rowan, M. J. & Anwyl, R. NMDA receptor- and metabotropic glutamate receptor-dependent synaptic plasticity induced by high frequency stimulation in the rat dentate gyrus *in vitro. J. Physiol. (Lond.)* **533,** 745–755 (2001).

46. Raber, J. *et al.* Hypothalamic-pituitary-adrenal function in *Apoe*$^{-/-}$ mice: possible role in behavioral and metabolic alterations. *J. Neurosci.* **20,** 2064–2071 (2000).

47. Raber, J., LeFevour, A., Buttini, M. & Mucke, L. Androgens protect against Apolipoprotein E4-induced cognitive deficits. *J. Neurosci.* **22,** 5204–5209 (2002).

48. Dere, E., Huston, J. P. & De Souza Silva, M. A. Episodic-like memory in mice: simultaneous assessment of object, place and temporal order memory. *Brain Res. Protoc.* **16,** 10–19 (2005).

49. Benice, T., Rizk, A., Kohama, S., Pfankuch, T. & Raber, J. Sex-differences in age-related cognitive decline in C57BL/6J mice associated with increased brain microtubule-associated protein 2 and synaptophysin immunoreactivity. *Neuroscience* **137,** 413–423 (2006).

50. Johnson-Wood, K. *et al.* Amyloid precursor protein processing and Aβ$_{42}$ deposition in a transgenic mouse model of Alzheimer disease. *Proc. Natl Acad. Sci. USA* **94,** 1550–1555 (1997).

Author Contributions M.C. and L.M. conceptualized the study. M.C., B.H., J.H. and N.D. performed experiments, and all authors participated in designing experiments and in analysing and interpreting data. M.C., B.H. and L.M. wrote the manuscript. L.M. supervised the project.

ARTICLE

A selective role for dopamine in stimulus–reward learning

Shelly B. Flagel[1]*, Jeremy J. Clark[2]*, Terry E. Robinson[3], Leah Mayo[1], Alayna Czuj[3], Ingo Willuhn[2], Christina A. Akers[2], Sarah M. Clinton[1], Paul E. M. Phillips[2] & Huda Akil[1]

Individuals make choices and prioritize goals using complex processes that assign value to rewards and associated stimuli. During Pavlovian learning, previously neutral stimuli that predict rewards can acquire motivational properties, becoming attractive and desirable incentive stimuli. However, whether a cue acts solely as a predictor of reward, or also serves as an incentive stimulus, differs between individuals. Thus, individuals vary in the degree to which cues bias choice and potentially promote maladaptive behaviour. Here we use rats that differ in the incentive motivational properties they attribute to food cues to probe the role of the neurotransmitter dopamine in stimulus–reward learning. We show that intact dopamine transmission is not required for all forms of learning in which reward cues become effective predictors. Rather, dopamine acts selectively in a form of stimulus–reward learning in which incentive salience is assigned to reward cues. In individuals with a propensity for this form of learning, reward cues come to powerfully motivate and control behaviour. This work provides insight into the neurobiology of a form of stimulus–reward learning that confers increased susceptibility to disorders of impulse control.

Dopamine is central for reward-related processes[1,2], but the exact nature of its role remains controversial. Phasic neurotransmission in the mesolimbic dopamine system is initially triggered by the receipt of reward (unconditional stimulus, US), but shifts to a cue that predicts a reward (conditional stimulus, CS) after associative learning[3,4]. Dopamine responsiveness appears to encode discrepancies between rewards received and those predicted, consistent with a 'prediction error' teaching signal used in formal models of reinforcement learning[5,6]. Therefore, a popular hypothesis is that dopamine is used to update the predictive value of stimuli during associative learning[7]. In contrast, others have argued that the role of dopamine in reward is in attributing Pavlovian incentive value to cues that signal reward, rendering them desirable in their own right[8–11], and thereby increasing the pool of positive stimuli that have motivational control over behaviour. Until now it has been difficult to determine whether dopamine mediates the predictive or the motivational properties of reward-associated cues, because these two features are often acquired together. However, the extent to which a predictor of reward acquires incentive value differs between individuals, providing the opportunity to parse the role of dopamine in stimulus–reward learning.

Individual variation in behavioural responses to reward-associated stimuli can be seen using one of the simplest reward paradigms, Pavlovian conditioning. If a CS is presented immediately before US delivery at a separate location, some animals approach and engage the CS itself and go to the location of food delivery only upon CS termination. This conditional response (CR), which is maintained by Pavlovian contingency[12], is called 'sign-tracking' because animals are attracted to the cue or sign that indicates impending reward delivery. However, other individuals do not approach the CS, but during its presentation engage the location of US delivery, even though the US is not available until CS termination. This CR is called 'goal-tracking'[13]. The CS is an effective predictor in animals that learn either a sign-tracking or a goal-tracking response; it acts as an excitor, evoking a CR

in both. However, only in sign-trackers is the CS an attractive incentive stimulus, and only in sign-trackers is it strongly desired (that is, 'wanted'), in the sense that animals will work avidly to get it[14]. In rats selectively bred for differences in locomotor responses to a novel environment[15], high responders to novelty (bHR rats) consistently learn a sign-tracking CR but low responders to novelty (bLR rats) consistently learn a goal-tracking CR[16]. Here, we exploit these predictable phenotypes in the selectively bred rats, as well as normal variation in outbred rats, to probe the role of dopamine transmission in stimulus–reward learning in individuals that vary in the incentive value they assign to reward cues.

Stimulus–reward learning

bHR and bLR rats from the twentieth generation of selective breeding (S20) were used for behavioural analysis of Pavlovian conditional approach behaviour[16] (Fig. 1a–e). When presentation of a lever-CS was paired with food delivery both bHR and bLR rats developed a Pavlovian CR, but as we have described previously[16], the topography of the CR was different in the two groups. With training, bHR rats came to rapidly approach and engage the lever-CS (Fig. 1a, b), whereas upon CS presentation bLR rats came to rapidly approach and engage the location where food would be delivered (Fig. 1c and d; see detailed statistics in Supplementary Information). Both bHR and bLR rats acquired their respective CRs as a function of training, given that there was a significant effect of number of sessions for all measures of sign-tracking behaviour for bHR rats (Fig. 1a, b; $P \leq 0.0001$), and of goal-tracking behaviour for bLR rats (Fig. 1c, d; $P \leq 0.0001$). Furthermore, bHR and bLR rats learned their respective CRs at the same rate, as indicated by analyses of variance in which session was treated as a continuous variable and the phenotypes were directly compared. There were non-significant phenotype × session interactions for (1) the number of contacts with the lever-CS for bHR rats versus the food-tray for bLR rats ($F_{(1, 236)} = 3.02$, $P = 0.08$) and (2) the latency to approach the lever-CS for bHR rats versus the

[1]Molecular and Behavioral Neuroscience Institute, University of Michigan, Michigan, USA. [2]Department of Psychiatry and Behavioral Sciences and Department of Pharmacology, University of Washington, Washington, USA. [3]Department of Psychology, University of Michigan, Michigan, USA.
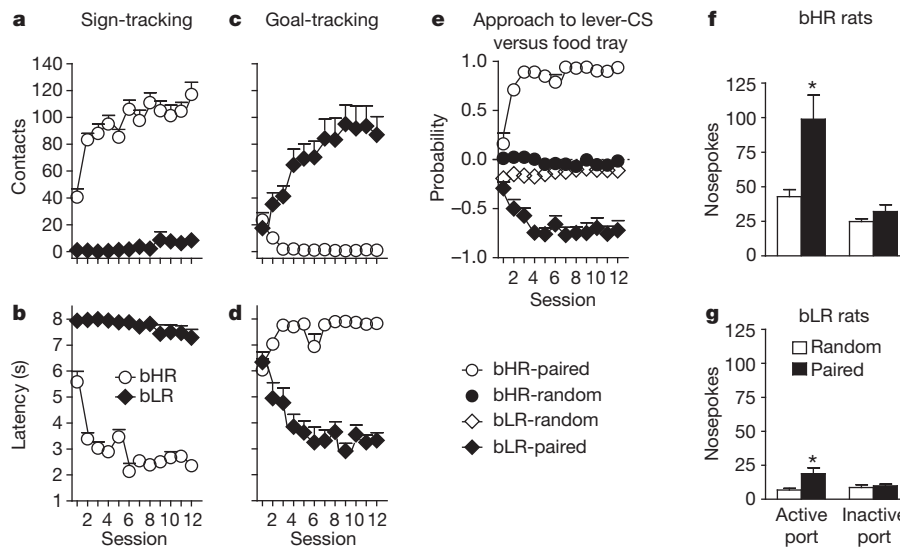*These authors contributed equally to this manuscript.

Figure 1 | Development of sign-tracking versus goal-tracking CRs in bHR and bLR rats. Behaviour directed towards the lever-CS (sign-tracking) is shown in **a** and **b** and behaviour directed towards the food-tray (goal-tracking) is shown in **c** and **d** (*n* = 10 per group). Data are shown as mean + s.e.m. **a**, Number of lever-CS contacts made during the 8-s CS period. **b**, Latency to the first lever-CS contact. **c**, Number of food-tray beam breaks during lever-CS presentation. **d**, Latency to the first beam break in the food-tray during lever-CS presentation. For all of these measures (**a–d**) there was a significant effect of phenotype, session, and a phenotype × session interaction ($P \leq 0.0001$). **e**, Probability of

approach to the lever minus the probability of approach to the food-tray shown as mean+ s.e.m. A score of zero indicates that neither approach to the lever-CS nor approach to the food-tray was dominant. **f**, **g**, Test for conditioned reinforcement illustrated as the number (mean + s.e.m.) of active and inactive nosepokes in bred rats that received either paired (bHR rats, *n* = 10; bLR rats, *n* = 9) or pseudorandom (bHR rats, *n* = 9; bLR rats, *n* = 9) CS–US presentations. Rats in the paired groups poked more in the active port than did random groups of the same phenotype (*$P < 0.02$), but the magnitude of this effect was greater for bHR rats (phenotype × group interaction, $P = 0.04$).

food-tray for bLR rats ($F_{(1, 236)} = 0.93$, $P = 0.34$). Importantly, rats that received non-contiguous (pseudorandom) presentations of the CS and the US did not learn either a sign-tracking or a goal-tracking CR (Fig. 1e).

These data indicate that the CS acquired one defining property of an incentive stimulus in bHR rats but not bLR rats: the ability to attract. Another feature of an incentive stimulus is to be 'wanted' and as such animals should work to obtain it[10,17]. Therefore, we quantified the ability of the lever-CS to serve as a conditioned reinforcer in the two groups (Fig. 1f, g) in the absence of the food-US. Following Pavlovian training, rats were given the opportunity to perform an instrumental response (a nosepoke) for presentation of the lever-CS. Responses into a port designated 'active' resulted in the brief presentation of the lever-CS and responses into an 'inactive' port were without consequence. Both conditioned bHR and bLR rats made more active than inactive nose pokes, and more active nose pokes than control groups that received pseudorandom presentations of the CS and the US (Fig. 1f, g; detailed statistics in Supplementary Information). However, the lever-CS was a more effective conditioned reinforcer in bHR rats than in bLR rats, as indicated by a significant phenotype × group interaction for active nose pokes ($F_{(1, 33)} = 4.82$, $P = 0.04$), which controls for basal differences in nose-poke responding. Moreover, in outbred rats in which this baseline difference in responding does not exist, we have found similar results, indicating that the lever-CS is a more effective conditioned reinforcer for sign-trackers than goal-trackers[14]. In summary, the lever-CS was equally predictive, evoking a CR in both groups, but it acquired two properties of an incentive stimulus to a greater degree in bHR rats than bLR rats: it was more attractive, as indicated by approach behaviour (Fig. 1a) and more desirable, as indicated by its ability to serve as a conditioned reinforcer (Fig. 1f, g).

Dopamine signalling during stimulus–reward learning
The core of the nucleus accumbens is an important anatomical substrate for motivated behaviour[18,19] and has been specifically implicated as a site where dopamine acts to mediate the acquisition and/or performance of Pavlovian conditional approach behaviour[20–23]. Therefore,

we used fast-scan cyclic voltammetry (FSCV) at carbon-fibre microelectrodes[24] to characterize the pattern of phasic dopamine signalling in this region during Pavlovian conditioning (see Supplementary Fig. 1 for recording locations). Similarly to surgically naive animals, bHR rats learned a sign-tracking CR (session effect on lever contacts: $F_{(5, 20)} = 5.76$, $P = 0.002$) and bLR rats learned a goal-tracking CR (session effect on food-receptacle contacts: $F_{(5, 20)} = 5.18$, $P = 0.003$) during neurochemical data collection (Supplementary Fig. 2). Changes in latency during learning were very similar in each group for their respective CRs (main effect of session: $F_{(5, 40)} = 10.5$, $P < 0.0001$; main effect of phenotype: $F_{(1, 8)} = 0.13$, $P = 0.73$; session × phenotype interaction: $F_{(5, 40)} = 1.16$, $P = 0.35$), indicating that the CS acts as an equivalent predictor of reward in both groups. Therefore, if CS-evoked dopamine release encodes the strength of the reward prediction, as previously postulated[5–7], it should increase to a similar degree in both groups during learning; however, if it encodes the attribution of incentive value to the CS, then it should increase to a greater degree in sign-trackers than in goal-trackers. During the acquisition of conditional approach, CS-evoked dopamine release (Fig. 2 and Supplementary Fig. 3) increased in bHR rats relative to unpaired controls (pairing × session interaction: $F_{(5, 35)} = 4.58$, $P = 0.003$), but there was no such effect in bLR rats (Fig. 2 and Supplementary Fig. 3; pairing × session interaction: $F_{(5, 35)} = 0.94$, $P = 0.46$). Indeed, the trial-by-trial correlation between CS-evoked dopamine release and trial number was significant for bHR rats ($r^2 = 0.14$, $P < 0.0001$) but not bLR rats ($r^2 = 0.003$, $P = 0.54$), producing significantly different slopes ($P = 0.005$) and higher CS-evoked dopamine release in bHR rats after acquisition (Supplementary Fig. 4, session 6; $P = 0.04$). US-evoked dopamine release also differed between bHR and bLR rats during training (session × phenotype interaction: $F_{(5, 40)} = 6.09$, $P = 0.0003$), but for this stimulus dopamine release was lower after acquisition in bHR rats (session 6; $P = 0.002$; Supplementary Fig. 4). Collectively, these data highlight that bHR and bLR rats produce fundamentally different patterns of dopamine release in response to reward-related stimuli during learning (see Supplementary Videos 1 and 2). The CS and US signals diverge in bHR rats (stimulus × session interaction: $F_{(5, 40)} = 5.47$,
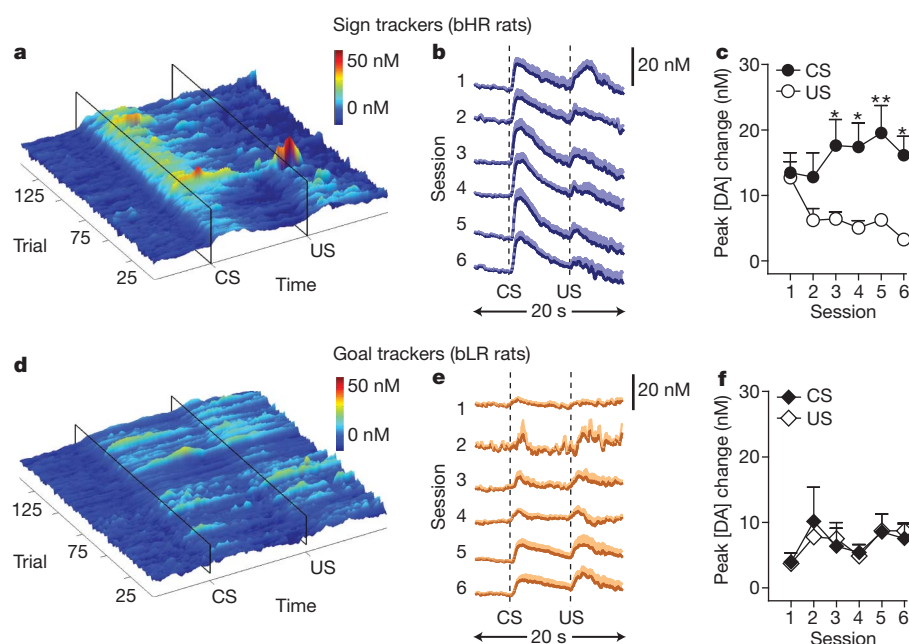
Figure 2 | Phasic dopamine signalling in response to CS and US presentation during the acquisition of Pavlovian conditional approach behaviour in bHR and bLR rats. Phasic dopamine release was recorded in the core of the nucleus accumbens using FSCV over six days of training. **a, d,** Representative surface plots depict trial-by-trial fluctuations in dopamine concentration [DA] during the twenty-second period around CS and US presentation in individual animals throughout training. **b, e,** Change in dopamine concentration (mean + s.e.m.) in response to CS and US presentation for each session of conditioning. **c, f,** Change in peak amplitude (mean + s.e.m.) of the dopamine signal observed in response to CS and US presentation for each session of conditioning ($n = 5$ per group; Bonferroni post-hoc comparison between CS- and US-evoked dopamine release: $*P < 0.05$; $**P < 0.01$). Panels **a–c** demonstrate that bHR rats, which developed a sign-tracking CR, show increasing phasic dopamine responses to CS presentation and decreasing responses to US presentation across the six sessions of training. In contrast, panels **d–f** demonstrate that bLR rats, which developed a goal-tracking CR, maintain phasic responses to US presentation throughout training.

$P = 0.0006$; Fig. 2c) but not bLR rats (stimulus × session interaction: $F_{(5, 40)} = 0.28$, $P = 0.92$; Fig. 2f).

Importantly, experiments conducted in commercially obtained outbred rats reproduced the pattern of dopamine release observed in the selectively bred rats (Fig. 3 and Supplementary Fig. 4). Specifically, there was an increase in CS-evoked and a decrease in US-evoked dopamine release during learning in outbred rats that learned a sign-tracking CR (stimulus × session interaction: $F_{(5, 50)} = 4.43$, $P = 0.002$; Fig. 3d), but not in those that learned a goal-tracking CR (stimulus × session interaction: $F_{(5, 40)} = 0.48$, $P = 0.72$; Fig. 3f). To test the robustness of these patterns of dopamine release, a subset of outbred rats received extended training. During four additional sessions, the profound differences in dopamine release between sign- and goal-trackers were stable (Supplementary Fig. 5), demonstrating that these differences are not limited to the initial stages of learning. The consistency of these dopamine patterns in selectively bred and outbred rats indicates that they are neurochemical signatures for sign- and goal-trackers rather than an artefact of selective breeding.

Stimulus–reward learning under dopamine blockade

Given the disparate patterns of dopamine signalling observed during learning a sign- versus goal-tracking CR, we tested whether the acquisition and performance of these CRs were differentially dependent on dopamine transmission. Systemic administration of flupenthixol, a nonspecific dopamine receptor antagonist, attenuated performance of the CR for both bHR and bLR rats. This effect was clearly evident when the antagonist was administered during training (Fig. 4, sessions 1–7). It was also observed after the rats had already acquired their respective CR (Supplementary Fig. 6), but this latter finding needs to be interpreted cautiously because of a non-specific effect on activity (Supplementary Fig. 6e). More importantly, when examined off flupenthixol during the eighth test session, bHR rats still failed to demonstrate a sign-tracking CR ($P \le 0.01$ versus saline, session 8; Fig. 4a–c), indicating that dopamine is necessary for both the performance and the learning of a sign-tracking CR, consistent with previous findings[21]. In contrast, flupenthixol had no effect on learning the CS–US association that lead to a goal-tracking CR ($P \ge 0.6$ versus saline, session 8; Fig. 4d–f), because on the drug-free session bLR rats showed a fully developed goal-tracking CR—their session 8 performance differed significantly from their session 1 performance ($P \le 0.0002$). Further, they differed from the bLR saline group on session 1 ($P \le 0.0001$), but did not differ from the bLR saline group on session 8. Thus, whereas dopamine may be

necessary for the performance of both sign-tracking and goal-tracking CRs, it is only necessary for acquisition of a sign-tracking CR, indicating that these forms of learning are mediated by distinct neural systems.

Collectively, these data provide several lines of evidence demonstrating that dopamine does not act as a universal teaching signal in stimulus–reward learning, but selectively participates in a form of stimulus–reward learning whereby Pavlovian incentive value is attributed to a CS. First, US-evoked dopamine release in the nucleus accumbens decreased during training in sign-trackers, but not in goal-trackers. Thus, during the acquisition of a goal-tracking CR, there is not a dopamine-mediated prediction-error teaching signal because, by definition, prediction errors become smaller as delivered rewards become better predicted. Second, the CS evoked dopamine release in both sign- and goal-tracking rats, but this signal increased to a greater degree in sign-trackers, which attributed incentive salience to the CS. These data indicate that the strength of the CS–US association is reflected by dopamine release to the CS only in some forms of stimulus–reward learning. Third, bHR rats that underwent Pavlovian training in the presence of a dopamine receptor antagonist did not acquire a sign-tracking CR, consistent with previous reports[8]; however, dopamine antagonism had no effect on learning a goal-tracking CR in bLR rats. Thus, learning a goal-tracking CR does not require intact dopamine transmission, whereas learning a sign-tracking CR does.

The attribution of incentive salience is the product of previous experience (that is, learned associations) interacting with an individual's genetic propensity and neurobiological state[8,17,25–27]. The selectively bred rats used in this study have distinctive behavioural phenotypes, including greater behavioural disinhibition and reduced impulse control in bHR rats[16]. Moreover, in these lines, unlike in outbred rats[14,28], there is a strong correlation between locomotor response to novelty and the tendency to sign-track[16]. These behavioural phenotypes are accompanied by baseline differences in dopamine transmission, with bHR rats showing elevated sensitivity to dopamine agonists, increased proportion of striatal D2 receptors in a high-affinity state, greater frequency of spontaneous dopamine transients[16], and higher reward-related dopamine release before conditioning, all of which could enhance their attribution of incentive salience to reward cues[29,30]. However, basal differences in dopaminergic tone do not provide the full explanation for differences in learning styles and associated dopamine responsiveness. Outbred rats with similar baseline locomotor activity[14] and similar baseline levels of reward-related
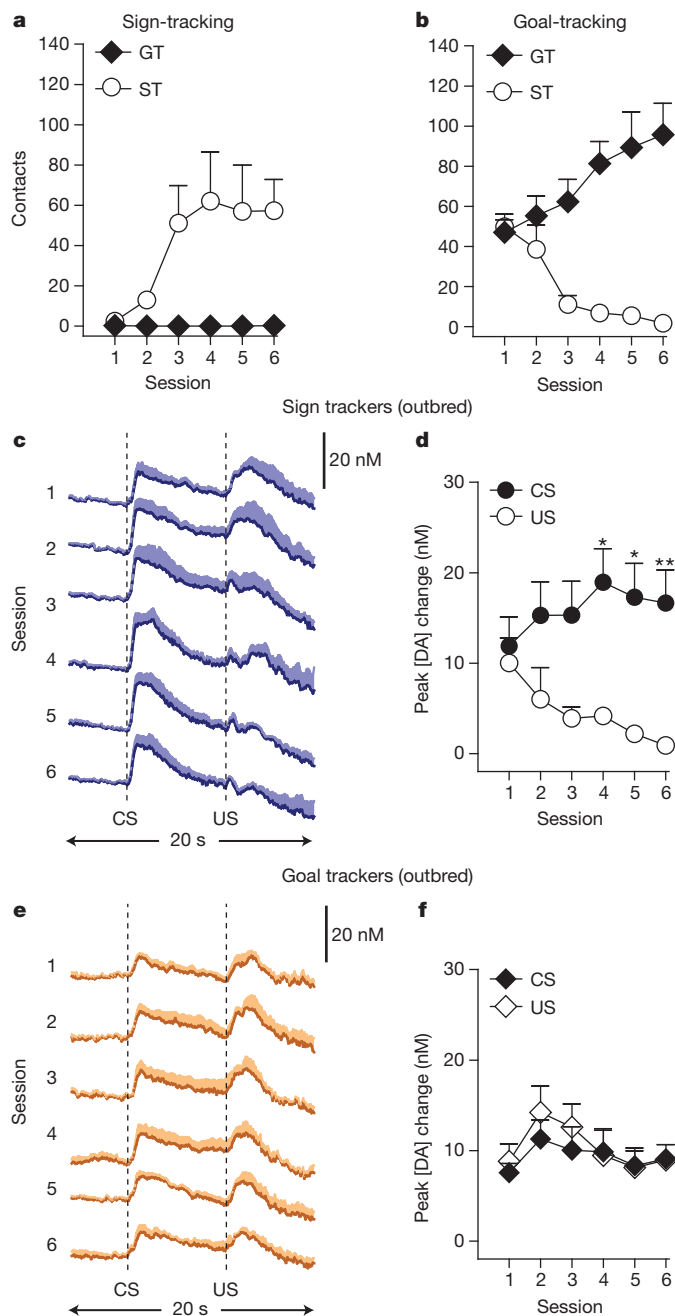
Figure 3 | Conditional responses and phasic dopamine signalling in response to CS and US presentation in outbred rats. Phasic dopamine release was recorded in the core of the nucleus accumbens using FSCV across six days of training. **a, b**, Behaviour directed towards the lever-CS (sign-tracking) (**a**) and behaviour directed towards the food-tray (goal-tracking) (**b**) during conditioning. Learning was evident in both groups because there was a significant effect of session both for rats that learned a sign-tracking response ($n = 6$; session effect on lever contacts: $F_{(5,25)} = 11.85$, $P = 0.0001$) and for rats that learned a goal-tracking response ($n = 5$; session effect on food-receptacle contacts: $F_{(5,20)} = 3.09$, $P = 0.03$). **c, e**, Change in dopamine concentration (mean + s.e.m.) in response to CS and US presentation for each session of conditioning. **d, f**, Change in peak amplitude (mean + s.e.m.) of the dopamine signal observed in response to CS and US presentation for each session of conditioning. (Bonferroni post-hoc comparison between CS- and US-evoked dopamine release: *$P < 0.05$; **$P < 0.01$). Panels **c** and **d** demonstrate that animals developing a sign-tracking CR ($n = 6$) show increasing phasic dopamine responses to CS presentation and decreasing responses to US presentation consistent with bHR rats. Panels **e–f** demonstrate that animals developing a goal-tracking CR ($n = 5$) maintain phasic responses to US presentation consistent with bLR rats.
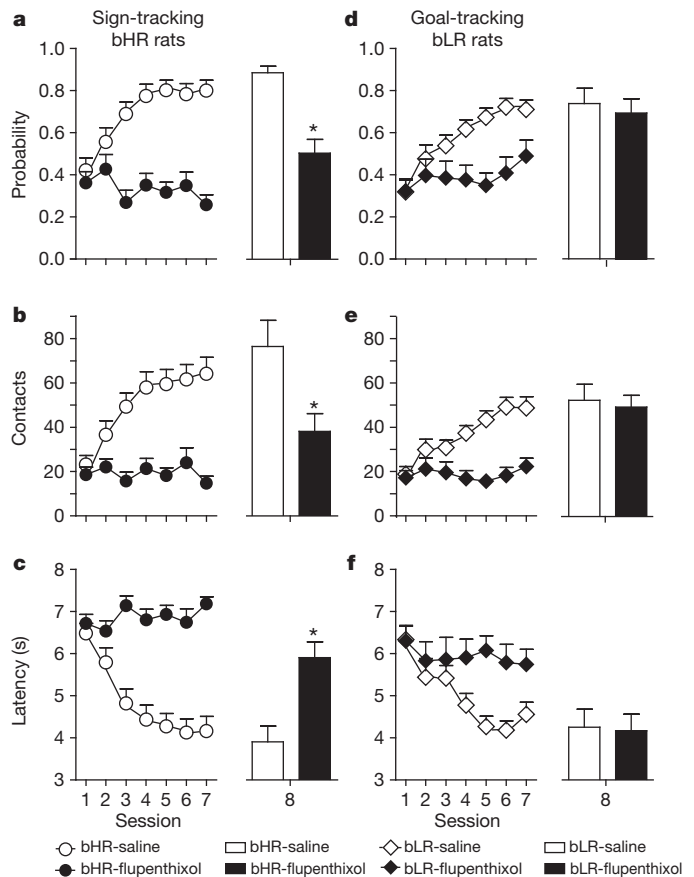
Figure 4 | Dopamine is necessary for learning CS–US associations that lead to sign-tracking, but not goal-tracking. a–c, The effects of flupenthixol on sign-tracking. **a**, Probability of approaching the lever-CS. **b**, Number of contacts with the lever-CS. **c**, Latency to contact the lever-CS. **d–f**, The effects of flupenthixol on goal-tracking. **d**, Probability of approaching the food-tray during lever-CS presentation. **e**, Number of contacts with the food-tray during lever-CS presentation. **f**, Latency to contact the food-tray during lever-CS presentation. Data are expressed as mean + s.e.m. Flupenthixol (sessions 1–7) blocked the performance of both sign-tracking and goal-tracking CRs. To determine whether flupenthixol influenced performance or learning of a CR, behaviour was examined following a saline injection on session 8 for all rats. bLR rats that were treated with flupenthixol before sessions 1–7 ($n = 16$) responded similarly to the bLR saline group ($n = 10$) on all measures of goal-tracking behaviour on session 8, whereas bHR rats treated with flupenthixol ($n = 22$) differed significantly from the bHR saline group ($n = 10$) on session 8 (*$P < 0.01$, saline versus flupenthixol). Thus, bLR rats learned the CS–US association that produced a goal-tracking CR even though the drug prevented the expression of this behaviour during training. Parenthetically, bHR rats treated with flupenthixol did not develop a goal-tracking CR.

dopamine release in the nucleus accumbens (see Fig. 3), differ in whether they are prone to learn a sign-tracking or goal-tracking CR, but they still develop patterns of dopamine release specific to that CR. Therefore, it appears that different mechanisms control basal dopamine neurotransmission versus the unique pattern of dopamine responsiveness to a reward cue.

The neural mechanisms underlying sign- and goal-tracking behaviour remain to be elucidated. Here we have shown that stimulus–reward associations that produce different CRs are mediated by different neural circuitry. Previous research using site-specific dopamine antagonism[21] and dopamine-specific lesions[22] indicated that dopamine acts in the nucleus accumbens core to support the learning and performance of sign-tracking behaviour. This work demonstrates that dopamine-encoded prediction-error signals are indeed present in the nucleus accumbens of sign-trackers, but not in the nucleus accumbens of goal-trackers. Although these neurochemical data alone do not rule out the possibility that prediction-error signals are present in other

dopamine terminal regions, the results from systemic dopamine antagonism demonstrate that intact dopamine transmission is generally not required for learning of a goal-tracking CR.

We thus show that dopamine is an integral part of stimulus–reward learning that is specifically associated with the attribution of incentive salience to reward cues. Individuals who attribute reward cues with incentive salience find it more difficult to resist such cues, a feature associated with reduced impulse control[16,31]. Human motivated behaviour is subject to a wide span of individual differences ranging from highly deliberative to highly impulsive actions directed towards the acquisition of rewards[32]. This work provides insight into the biological basis of these individual differences, and may provide an important step for understanding and treating impulse-control problems that are prevalent across several psychiatric disorders.

METHODS SUMMARY

The majority of these studies were conducted with adult male Sprague–Dawley rats from a selective-breeding colony which has been previously described[15]. The data presented here were obtained from bHR and bLR rats from generations S18–S22. Equipment and procedures for Pavlovian conditioning have been described in detail elsewhere[14,16]. Selectively bred rats from generations S18, S20 and S21 were transported from the University of Michigan to the University of Washington for the FSCV experiments. During each behaviour session, chronically implanted microsensors, placed in the core of the nucleus accumbens, were connected to a head-mounted voltammetric amplifier for detection of dopamine by FSCV[24]. Voltammetric scans were repeated every 100 ms to obtain a sampling rate of 10 Hz. Voltammetric analysis was carried out using software written in LabVIEW (National Instruments). On completion of the FSCV experiments, recording sites were verified using standard histological procedures. To examine the effects of flupenthixol (Sigma; dissolved in 0.9% NaCl) on the performance of sign-tracking and goal-tracking behaviour, rats received an injection (intraperitoneal, i.p.) of 150, 300 or 600 μg kg^{-1} of the drug one hour before Pavlovian conditioning sessions 9, 11 and 13. Doses of the drug were counterbalanced between groups and interspersed with saline injections (i.p., 0.9% NaCl; before sessions 8, 10, 12 and 14) to prevent any cumulative drug effects. To examine the effects of flupenthixol on the acquisition of sign-tracking and goal-tracking behaviour, rats received an injection (i.p.) of either saline or 225 μg kg^{-1} of the drug one hour before Pavlovian conditioning sessions 1–7.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Schultz, W. Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* **57,** 87–115 (2006).
2. Wise, R. A. Dopamine, learning and motivation. *Nature Rev. Neurosci.* **5,** 483–494 (2004).
3. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neurosci.* **10,** 1020–1028 (2007).
4. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
5. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16,** 1936–1947 (1996).
6. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412,** 43–48 (2001).
7. Balleine, B. W., Daw, N. D. & O'Doherty, J. P. in *Neuroeconomics: Decision Making and the Brain* (eds Glimcher, P. W., Camerer, C. F., Fehr, E. & Poldrack, R. A.) 367–389 (Academic Press, 2008).
8. Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* **191,** 391–431 (2007).
9. Berridge, K. C. & Robinson, T. E. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* **28,** 309–369 (1998).
10. Berridge, K. C., Robinson, T. E. & Aldridge, J. W. Dissecting components of reward: 'liking', 'wanting', and learning. *Curr. Opin. Pharmacol.* **9,** 65–73 (2009).
11. Panksepp, J. Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* **14,** 30–80 (2005).
12. Hearst, E. & Jenkins, H. *Sign-Tracking: The Stimulus-Reinforcer Relation and Directed Action* (Monograph of the Psychonomic Society, 1974).
13. Boakes, R. in *Operant-Pavlovian Interactions* (eds Davis, H. & Hurwitz, H. M. B.) 67–97 (Erlbaum, 1977).
14. Robinson, T. E. & Flagel, S. B. Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. *Biol. Psychiat.* **65,** 869–873 (2009).
15. Stead, J. D. *et al.* Selective breeding for divergence in novelty-seeking traits: heritability and enrichment in spontaneous anxiety-related behaviors. *Behav. Genet.* **36,** 697–712 (2006).
16. Flagel, S. B. *et al.* An animal model of genetic vulnerability to behavioral disinhibition and responsiveness to reward-related cues: implications for addiction. *Neuropsychopharmacology* **35,** 388–400 (2010).
17. Berridge, K. C. in *Psychology of Learning and Motivation* (ed. Medin, D. L.) 223–278 (Academic Press, 2001).
18. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26,** 321–352 (2002).
19. Kelley, A. E. Functional specificity of ventral striatal compartments in appetitive behaviors. *Ann. NY Acad. Sci.* **877,** 71–90 (1999).
20. Dalley, J. W. *et al.* Time-limited modulation of appetitive Pavlovian memory by D1 and NMDA receptors in the nucleus accumbens. *Proc. Natl Acad. Sci. USA* **102,** 6189–6194 (2005).
21. Di Ciano, P., Cardinal, R. N., Cowell, R. A., Little, S. J. & Everitt, B. J. Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *J. Neurosci.* **21,** 9471–9477 (2001).
22. Parkinson, J. A. *et al.* Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behav. Brain Res.* **137,** 149–163 (2002).
23. Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W. & Everitt, B. J. Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. *J. Neurosci.* **19,** 2401–2411 (1999).
24. Clark, J. J. *et al.* Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. *Nature Methods* **7,** 126–129 (2010).
25. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Brain Res. Rev.* **18,** 247–291 (1993).
26. Tindell, A. J., Smith, K. S., Berridge, K. C. & Aldridge, J. W. Dynamic computation of incentive salience: "wanting" what was never "liked". *J. Neurosci.* **29,** 12220–12228 (2009).
27. Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S. & Aldridge, J. W. A neural computational model of incentive salience. *PLOS Comput. Biol.* **5,** e1000437 (2009).
28. Beckmann, J. S., Marusich, J. A., Gipson, C. D. & Bardo, M. T. Novelty seeking, incentive salience and acquisition of cocaine self-administration in the rat. *Behav. Brain Res.* **216,** 159–165 (2011).
29. Wyvell, C. L. & Berridge, K. C. Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. *J. Neurosci.* **20,** 8122–8130 (2000).
30. Wyvell, C. L. & Berridge, K. C. Incentive sensitization by previous amphetamine exposure: increased cue-triggered "wanting" for sucrose reward. *J. Neurosci.* **21,** 7831–7840 (2001).
31. Tomie, A., Aguado, A. S., Pohorecky, L. A. & Benjamin, D. Ethanol induces impulsive-like responding in a delay-of-reward operant choice procedure: impulsivity predicts autoshaping. *Psychopharmacology* **139,** 376–382 (1998).
32. Kuo, W. J., Sjostrom, T., Chen, Y. P., Wang, Y. H. & Huang, C. Y. Intuition and deliberation: two systems for strategizing in the brain. *Science* **324,** 519–522 (2009).

METHODS

Animals. Adult male Sprague-Dawley rats selectively bred for reactivity to a novel environment were used for the majority of these studies[15]. The data presented here were obtained from bHR and bLR rats from generations S18 to S22. The experiments followed the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioural Research (National Research Council 2003) and the procedures were approved by the University Committee on the Use and Care of Animals. Unless otherwise indicated, rats were housed in pairs and kept on a 12-h light/12-h dark cycle (lights on 06:00 h) with controlled temperature and humidity and food and water were available *ad libitum*.

Voltammetry studies were conducted at the University of Washington using bHR and bLR rats from generations S18, S20 and S21as well as male Sprague-Dawley rats obtained from Charles River weighing between 300 g and 350 g upon arrival. These rats were housed individually and kept on a 12-h light/12-h dark cycle (lights on at 0700) with controlled temperature and humidity. Prior to behavioural training, food was restricted so that rats maintained 90% of their free-feeding body weight and water was available *ad libitum*. All animal procedures followed the University of Washington Institutional Animal Care and Use Committee guidelines.

Screening for selectively bred phenotypes. To confirm the selectively bred phenotypes, each generation of rats were screened for locomotor activity in novel test chambers at around 60 days of age, as previously described[15,33].

Pavlovian conditioning procedures. Equipment and procedures for Pavlovian conditioning have been described in detail elsewhere[14,16]. Briefly, standard Med Associates test chambers were equipped with a food-tray located in the middle of the front wall and a retractable lever located to the left or right of the food-tray (counterbalanced). The lever required only a 10-g force to operate, such that most contacts with the lever were detected and recorded as a 'lever press'. Operation of the pellet dispenser (Med Associates) delivered one 45-mg banana-flavoured food pellet (Bio-Serv) into the food-tray. Head entries into the food-tray were recorded each time the rat broke a photobeam located inside the receptacle.

All Pavlovian training sessions were conducted between 13:00 h and 18:00 h. Banana-flavoured food pellets were placed into the rats' home cages for 2 days before training to familiarize the animals with this food (the unconditioned stimulus, US). Two pre-training sessions were conducted that consisted of the delivery of 50 food pellets, which were randomly delivered on a variable-interval 30-s schedule (25-min session), during which it was determined whether the rats were reliably retrieving the food pellets. Following pre-training sessions, Pavlovian training sessions consisted of the presentation of the illuminated lever (conditioned stimulus, CS) in the chamber for 8 s, and then immediately upon its retraction a 45-mg food pellet (US) was delivered into the food-tray (the 'goal'). The CS was presented on a random-interval 90-s schedule and each Pavlovian training session consisted of 25 trials (or CS–US pairings). Training continued for 6–12 sessions. Rats in the 'random' groups received presentations of the CS and US, each on a variable-interval 90-s schedule.

The following events were recorded using Med Associates software: (1) the number of lever-CS contacts, (2) the latency to the first lever-CS contact, (3) the number of food-tray entries during lever-CS presentation, and (4) the latency to the first food-tray entry during lever-CS presentation. It is important to note that no response is required for the rat to receive the reward (US), yet distinct CRs emerge as a result of Pavlovian conditioning. The outcome measures listed above allow us to examine CS-directed (sign-tracking) versus goal-directed (goal-tracking) responses. Using these measures we calculated the probability that a rat would approach the lever-CS or the food-tray as well as the difference in its probability of approaching the lever-CS versus the food-tray.

Statistical analysis of Pavlovian conditional responses. Differences in the conditional response that emerged across training sessions were analysed using linear mixed effects models (SPSS 17.0; see also ref. 34), in which phenotype and session were treated as independent variables. In addition, the effect of session for each phenotype was analysed separately. For all analyses, the covariance structure was explored and modelled appropriately. When significant main effects or interactions were detected, Bonferroni post-hoc comparisons were made. The differences in the probability of approaching the lever-CS versus the food-tray (Fig. 1e) were further examined using one-sample *t*-tests (with hypothesized value of 0) to determine whether either phenotype exhibited a preference for the lever-CS or the food-tray.

Conditioned reinforcement test. The conditioned reinforcement test occurred one day after the last of 12 Pavlovian training sessions. The conditioned reinforcement test was conducted in the same standard Med Associates chambers as described above. However, for the purposes of this test the chambers were rearranged such that the retractable lever was placed in the centre of the front wall in between two nosepoke ports. The 'active' port was placed on the side of the wall opposite to the location of the lever-CS during Pavlovian training. During the 40-min conditioned reinforcement test nosepokes into the port designated 'active' resulted in the 2-s presentation of the illuminated lever, whereas pokes into the other 'inactive' port were without consequence. The number of nosepokes into the active and inactive ports and the number of contacts with the lever were recorded throughout the test session.

Statistical analysis of conditioned reinforcement. Performance on the conditioned reinforcement test was analysed using a three-way analysis of variance (ANOVA) in which phenotype, group (paired versus unpaired) and port (active versus inactive) were treated as independent variables and the number of pokes as the dependent variable. Further analyses were then conducted to determine the effect of group or port for each phenotype and the effect of phenotype or group for each port.

FSCV. The following procedures were in accordance with the University of Washington Institutional Animal Care and Use Committee guidelines. Surgical preparation for *in vivo* voltammetry used an aseptic technique. Rats were anaesthetized with isofluorane and placed in a stereotaxic frame. The scalp was swabbed with 10% povidone iodine, bathed with a mixture of lidocaine (0.5 mg kg^{-1}) and bupivicaine (0.5 mg kg^{-1}), and incised to expose the cranium. Holes were drilled and cleared of dura mater above the nucleus accumbens core (1.3-mm lateral and 1.3-mm rostral from the bregma), and at convenient locations for a reference electrode and three anchor screws. The reference electrode and anchor screws were positioned and secured with cranioplastic cement, leaving the working electrode holes exposed. Once the cement cured, the microsensors were attached to the voltammetric amplifier and lowered into the target recording regions (the core of the nucleus accumbens, 7.0-mm ventral of dura mater). Finally, cranioplastic cement was applied to the part of the cranium still exposed to secure the working electrode.

Voltammetric measurement. During all experimental sessions, chronically implanted microsensors were connected to a head-mounted voltammetric amplifier for dopamine detection by FSCV[24]. Voltammetric scans were repeated every 100 ms to obtain a sampling rate of 10 Hz. When dopamine is present at the surface of the electrode during a voltammetric scan, it is oxidized during the anodic sweep to form dopamine-*o*-quinone (peak reaction at approximately +0.7 V), which is reduced back to dopamine in the cathodic sweep (peak reaction at approximately −0.3 V). The ensuing flux of electrons is measured as current and is directly proportional to the number of molecules that undergo the electrolysis. The redox current obtained from each scan provides a chemical signature that is characteristic of the analyte, allowing resolution of dopamine from other substances. For quantification of changes in dopamine concentration over time, the current at its peak oxidation potential can be plotted for successive voltammetric scans. Waveform generation, data acquisition and analysis were carried out on a PC-based system using two PCI multifunction data acquisition cards and software written in LabVIEW (National Instruments).

Statistical analysis of voltammetry data. Voltammetric data analysis was carried out using software written in LabVIEW (National Instruments) and low-pass filtered at 2,000 Hz. Dopamine was isolated from the voltammetric signal with chemometric analysis[35] using a standard training set based upon stimulated dopamine release detected by chronically implanted electrodes. Dopamine concentration was estimated on the basis of the average post-implantation sensitivity of electrodes[24]. Before the generation of surface plots and analysis of peak values, all data were smoothed with a 5-point within-trial running average. Peak dopamine values in response to the US and CS were obtained by taking the largest value in the 3-s period after stimulus presentation. Peak values were then compared using mixed models ANOVA with training session as the repeated measure and stimulus (CS and US) or phenotype (bHR and bLR) as the between-group measure. Peak CS-evoked dopamine signalling was also analysed across trials using linear regression. The slopes obtained for the regression were compared between groups using independent, two-sample *t*-tests. All post-hoc comparisons were made with the Bonferroni correction for multiple tests. All statistical analyses were carried out using Prism (GraphPad Software). Voltammetric data for dopamine responses to the CS and US were also analysed using an area-under-the-curve approach. This approach did not alter the statistical effects of any comparison reported in the paper for peak dopamine value (specific statistical results not shown).

Histological verification of recording site. On completion of experimentation, animals were anesthetized with intraperitoneal ketamine (100 mg kg^{-1}) and xylazine (20 mg kg^{-1}) and then transcardially perfused with saline followed by 4% paraformaldehyde. Brains were removed and post-fixed in paraformaldehyde for 24 h and then rapidly frozen in an isopentane bath (~5 min), sliced on a cryostat (50-µm coronal sections, 20 °C) and stained with cresyl violet to aid in visualization of anatomical structures.

Effects of flupenthixol on sign-tracking and goal-tracking performance. The effects of flupenthixol (a D1/D2 antagonist; Sigma) on the performance of sign-tracking and goal-tracking behaviour were examined after seven sessions of Pavlovian conditioning. All rats received an injection (i.p.) of 150, 300 or

$600\,\mu g\,kg^{-1}$ of the drug one hour before Pavlovian conditioning sessions 9, 11 and 13. Doses of the drug (dissolved in 0.9% NaCl) were counterbalanced between groups and interspersed with saline injections (i.p., 0.9% NaCl; before sessions 8, 10, 12 and 14) to prevent any cumulative drug effects. The following measures were recorded to examine the effects of the drug on the CR: (1) the number of lever-CS contacts, (2) the latency to the first lever-CS contact, (3) the number of food-tray entries during lever-CS presentation, and (4) the latency to the first food-tray entry during lever-CS presentation. In addition, a nosepoke port was added to the test chamber on the wall opposite the retractable lever and responses into this port were recorded as an index of nonspecific activity. For all measures the response to saline was averaged (across sessions 8, 10, 12 and 14) and compared to the response following each of the three doses of flupenthixol.

Statistical analysis of effects of flupenthixol on performance of the CRs. The effects of flupenthixol on the performance of sign-tracking and goal-tracking behaviour (Supplementary Fig. 6) were analysed using linear mixed effects models with phenotype and dose treated as independent variables. Each phenotype was also analysed separately to determine the effect of dose on a given behaviour and Bonferroni post-hoc comparisons were made to determine whether behaviour at a given dose was significantly different from that in response to saline.

Effects of flupenthixol on the learning of sign-tracking and goal-tracking. The effects of flupenthixol on the acquisition of sign-tracking and goal-tracking CRs were examined using two generations of bred rats (S21 and S22). Rats received an injection of either saline (i.p.; 0.9% NaCl) or $225\,\mu g\,kg^{-1}$ of flupenthixol one hour before Pavlovian conditioning sessions 1–7. This dose of drug was chosen based on the 'performance' study described above, because we wanted to avoid any nonspecific inhibitory effects on motor activity. Rats from both generations that received flupenthixol before sessions 1–7 then received an injection of saline before session 8. However, only rats from the S22 generation that received saline

before sessions 1–7 also received saline before session 8. Thus, the number of rats that received saline during training and were also pretreated with saline before session 8 is lower than that for the other groups (that is, on session 8, bHR saline, $n = 10$; bLR saline, $n = 10$). The following measures were recorded and analysed to examine the effects of flupenthixol on sign-tracking and goal-tracking behaviour: (1) the number of lever-CS contacts, (2) the latency to the first lever-CS contact, (3) the number of food-tray entries during lever-CS presentation, and (4) the latency to the first food-tray entry during lever-CS presentation.

Statistical analysis of effects of flupenthixol on the learning of the CRs. Linear mixed effects models were used to examine the effects of flupenthixol on the performance and learning of sign-tracking or goal-tracking behaviour (Supplementary Fig. 6). For these analyses each phenotype was analysed separately to determine the effect of dose on a given behaviour and treatment (saline versus flupenthixol) and session (1–7) were treated as independent variables. To determine whether flupenthixol prevented the expression of the conditioned response or the learning of a conditioned response we also examined behaviour following a saline injection on session 8 (drug-free test session). Behaviour on session 8 was compared between treatment groups using an unpaired t-test for each phenotype separately. We also compared the response on session 8 of the groups that received flupenthixol during training to that of the group that received flupenthixol on session 1 (using a paired t-test) and to that of the saline control group on session 1 (using an unpaired t-test).

33. Clinton, S. M, *et al.* Individual differences in novelty-seeking and emotional reactivity correlate with variation in maternal behavior. *Horm. Behav.* **51,** 655–664 (2007).
34. Verbeke, G. & Molenberghs, G. *Linear Mixed Models for Longitudinal Data* (Springer, 2000).
35. Heien, M. L. Johnson, M. A. & Wightman, R. M. Resolving neurotransmitters detected by fast-scan cyclic voltammetry. *Anal. Chem.* **76,** 5697–5704 (2004).

ARTICLE

Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza

Fabienne Maillet[1]*, Véréna Poinsot[2]*, Olivier André[1]*, Virginie Puech–Pagès[3,4], Alexandra Haouy[1], Monique Gueunier[1,3,4], Laurence Cromer[1], Delphine Giraudet[1], Damien Formey[3,4], Andreas Niebel[1], Eduardo Andres Martinez[5], Hugues Driguez[5], Guillaume Bécard[3,4] & Jean Dénarié[1]

Arbuscular mycorrhiza (AM) is a root endosymbiosis between plants and glomeromycete fungi. It is the most widespread terrestrial plant symbiosis, improving plant uptake of water and mineral nutrients. Yet, despite its crucial role in land ecosystems, molecular mechanisms leading to its formation are just beginning to be unravelled. Recent evidence suggests that AM fungi produce diffusible symbiotic signals. Here we show that *Glomus intraradices* secretes symbiotic signals that are a mixture of sulphated and non–sulphated simple lipochitooligosaccharides (LCOs), which stimulate formation of AM in plant species of diverse families (Fabaceae, Asteraceae and Umbelliferae). In the legume *Medicago truncatula* these signals stimulate root growth and branching by the symbiotic DMI signalling pathway. These findings provide a better understanding of the evolution of signalling mechanisms involved in plant root endosymbioses and will greatly facilitate their molecular dissection. They also open the way to using these natural and very active molecules in agriculture.

Arbuscular mycorrhiza (AM) is a root endosymbiosis between fungi of the ancient phylum Glomeromycota and terrestrial plants. AM is formed by 70–90% of land plant species, improving the uptake of water and mineral nutrients[1]. Despite its crucial ecological importance, the mechanisms underlying the formation of this symbiosis are poorly understood. This is essentially because of the obligate biotrophy and multinucleate nature of AM fungi[2]. In contrast, the root endosymbiosis associating nitrogen-fixing rhizobia and legumes involves genetically amenable bacterial partners, which has led to a better understanding of the mechanisms involved in the development of this association[3–5]. Nod factors were shown to be LCO signals produced by most rhizobia and required for early steps of legume infection and root nodule organogenesis[6] (Supplementary Note 1). Legumes can establish symbioses both with rhizobia and with AM fungi, and the use of model legumes has facilitated the identification of plant genes essential for these associations[3,7,8]. In *M. truncatula* three of the genes that control the major Nod-factor signal-transduction pathway, *DMI1*, *DMI2* and *DMI3*, are required both for nodulation and for AM formation[9]. Thus it appears that these two different symbioses share central components of the signalling pathways used to trigger their symbiotic programmes[10–12]. A model was proposed in which rhizobial Nod factors and putative AM fungal signals, the so-called 'Myc factors', would activate signalling pathways having common components, the DMI proteins[7,9,13]. Recent findings indicate that Myc factors are diffusible compounds; spores separated by a membrane from roots of the host plant, or exudates of germinating spores, elicit plant symbiotic responses[14–17].

The AM symbiosis is extremely ancient and appeared more than 400 million years ago[7,18] whereas the rhizobium–legume symbiosis is estimated to have appeared about 60 million years ago[19]. It is thus probable that symbiotic signalling mechanisms evolved first in the AM symbiosis and were then recruited and adapted for the rhizobium–legume association[7]. We thus made the working hypothesis that Myc signals, which are produced by fungi able to synthesize chitin, are the

ancestors of the more recent Nod factors and are thus likely to be LCOs. Our strategy to purify them was based on the use of two bioassays that are very sensitive to a broad variety of LCOs: (1) the *Vicia sativa* root-hair branching assay (VsHab), which can detect several different non-sulphated LCOs[20]; (2) a transgenic line of *M. truncatula* carrying a fusion between the promoter region of the early nodulin gene *MtENOD11* and the *GUS* reporter gene (*ENOD11* assay[21]), which is induced by a variety of sulphated LCOs (F. Maillet, unpublished observations). In addition, we also used root branching of *M. truncatula* as a test for Myc signals[15].

G. intraradices secretes LCOs

Sterile exudates of carrot roots mycorrhized by *G. intraradices* were extracted with butanol and ethyl acetate. The butanol phase, being the most active on both *ENOD11* and VsHab bioassays, was further separated using reverse solid-phase extraction. The most active fraction, eluting with 50% acetonitrile (ACN), was then purified on semi-preparative reverse-phase high-performance liquid chromatography (HPLC). Two fractions with biological activity could be detected: fraction A (eluting at around 40% ACN) was active on *ENOD11*, and fraction B (eluting at around 60% ACN) was active on VsHab (Fig. 1a). Both fractions stimulated root branching in *M. truncatula*, as already shown for diffusible factors secreted by AM fungi[15] (Supplementary Fig. 1a). These fractions were submitted to ultraperformance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (UPLC/Q-ToF MS). Ion currents with mass to charge ratios (m/z) corresponding to hypothetical LCOs ranging from glucosamine dimers to hexamers with a large variety of possible N-substitutions (acyl and methyl) and O-substitutions (methyl, carbamoyl, acetyl, fucosyl, sulphate and so on), including those identified in rhizobial LCOs[6,22], were searched for in the chromatograms. Fraction A (F4 and F5; Fig. 1a) showed, in the negative mode electrospray ionization/mass spectrometry, 12 molecular ions that had m/z values of sulphated simple LCOs. Six molecular ions (m/z values of 1,105.5,

[1]Laboratoire des Interactions Plantes-Microorganismes, UMR 441/2594 INRA-CNRS, B.P. 52627, F-31326 Castanet-Tolosan CEDEX, France. [2]Laboratoire des IMRCP, UMR 5623 CNRS-Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse CEDEX, France. [3]Université de Toulouse, UPS, UMR 5546, Surfaces Cellulaires et Signalisation chez les Végétaux, BP 42617, F-31326 Castanet-Tolosan CEDEX, France. [4]CNRS, UMR 5546, BP 42617, F-31326 Castanet-Tolosan CEDEX, France. [5]Centre de Recherches sur les Macromolécules Végétales, CNRS (affiliated to Université Joseph Fourier), B.P. 53, F-38041 Grenoble CEDEX 9, France.
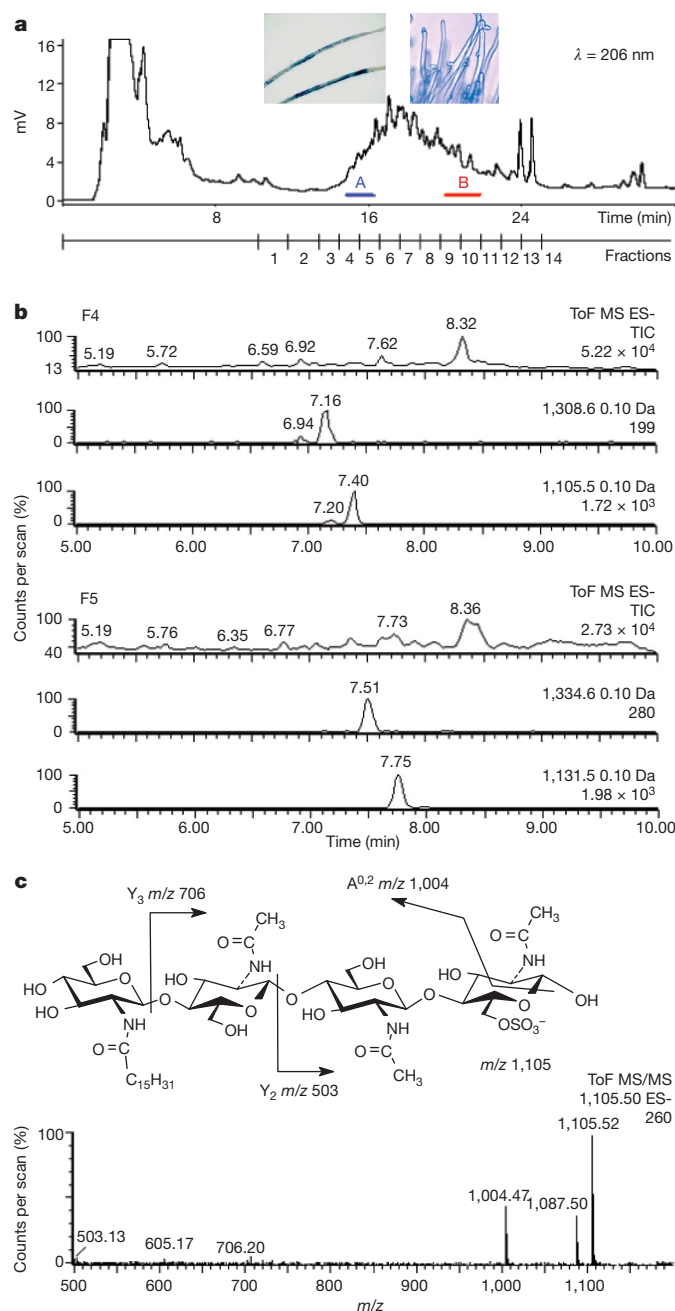*These authors contributed equally to this work.

Figure 1 | Detection and characterization of LCOs in mycorrhized carrot root exudates. a, Reverse-phase (C18) HPLC of exudate butanol extract separated two active fractions (test responses are given as inserts): A (fractions F4 and F5) and B (F9 and F10). **b**, UPLC/Q-ToF chromatograms of F4 and F5. Ion current m/z values of 1,308.6 and 1,105.5 ± 0.1 Da recorded for F4 are characteristic of LCO-V(C16:0, S) and LCO-IV(C16:0, S) respectively, whereas m/z values of 1,334.6 and 1,131.5 ± 0.1 Da for F5 are characteristic of LCO-V(C18:1, S) and LCO-IV(C18:1, S). **c**, Top, structure of LCO-IV(C16:0, S) showing Y and A fragmentations. Bottom, tandem mass spectrometry (MS/MS) spectrum of m/z value of 1,105.5. The presence of Y_3, Y_2 and $A^{0.2}$ fragments (in the correct ratios) and the parity change of successive fragments demonstrate the LCO nature of this compound.

1,103.5, 1,101.5, 1,133.5, 1,131.5, 1,129.5) could correspond to sulphated tetrameric LCOs (LCO-IV, S) N-acylated with a C16 or a C18 fatty acid with no, one or two unsaturations. Six other molecular ions (m/z values of 1,308.6, 1,306.6, 1,304.6, 1,336.6, 1,334.6 and 1,332.6) could correspond to sulphated pentameric LCOs (LCO-V, S) N-acylated with the same acyl chains (Fig. 1b and Supplementary Figs 2 and 3). Potential tetrameric LCOs were about tenfold more abundant than pentameric

ones. To confirm the structure of these molecular ions, UPLC/Q-ToF tandem MS was performed and the typical fragmentation of LCOs was observed (Fig. 1c, Supplementary Fig. 4 and Supplementary Note 2). Exact masses and isotopic profiles corresponded to the calculated ones (Supplementary Fig. 5). Fraction B (F9 and F10; Fig. 1a) was also analysed in UPLC/Q-ToF MS and tandem MS in the positive mode, and characteristic ion currents and B fragmentation, corresponding to non-sulphated LCO-IV(C16:0) and LCO-IV(C18:1), were observed (Supplementary Fig. 6). In fractions F1 to F3, we could not detect any ion currents corresponding to more hydrophilic LCOs. In fractions F6 to F8, we could not detect more hydrophobic LCOs. Thus a mixture of sulphated and non-sulphated simple LCOs is present in exudates of mycorrhized roots. In exudates of non-mycorrhized roots, grown in the same conditions but without AM fungi, no LCOs could be detected by very sensitive LC/Q-Trap analyses, supporting the fungal origin of these LCOs (Supplementary Fig. 7). To confirm this, sterile germinating spore exudates (GSE) of *G. intraradices* were analysed. A GSE butanol extract elicited *M. truncatula* root branching (Supplementary Fig. 1b). Extracts were analysed by LC/Q-Trap in the positive multiple reaction monitoring (MRM) mode. The MRM traces, characteristic of LCO-IV(C18:1) (1,053 > 426, 1,053 > 629, 1,053 > 832; Fig. 2c) and LCO-IV(C16:0) (1,027 > 400, 1,027 > 603, 1,027 > 806, data not shown) were observed in fraction B (Fig. 2a). HPLC retention times and MRM traces observed in GSE were identical to those of corresponding synthetic Myc-LCOs (see Supplementary Note 3 and Fig. 3c), demonstrating the presence of LCOs in fungal exudates. Biological activity of spore exudates increased during the 9-day germination period as well as quantities of LCOs detected by MRM mass spectrometry showing that bioactive LCOs are secreted *de novo* by germinating spores (data not shown). Similar analyses of fraction A showed traces of sulphated LCOs (Fig. 2d). The presence of sulphated LCOs in GSE was confirmed by mild methanolysis of the HPLC fraction A. This mild methanolysis, which affects the sulphate group but not the saccharidic bonds of LCOs[23], resulted in a clear shift in the biological activity of fraction A, with a strong decrease in the *ENOD11* response (characteristic of sulphated LCOs) to an increase in the VsHab response (characteristic of non-sulphated LCOs) (Supplementary Fig. 8).

Palmitic acid (C16:0) and a C18:1 fatty acid are the major N-acyl substitutions of AM fungal LCOs. To determine the structure of the C18:1 substituent, fatty acids resulting from hydrolysis of the HPLC fraction B from GSE were analysed using gas chromatography/ positive electron impact mass spectrometry. Comparison of retention times and spectra with those of commercial standards demonstrated that the major C18:1 fatty acid present in the fraction was oleic acid (C18:1 Δ9Z) (Supplementary Note 2). The spectrum exhibited 97% quality compared with the standard (Fig. 2e). We can conclude that the AM fungus *G. intraradices* secretes a mixture of sulphated and non-sulphated simple LCOs. The proportion of sulphated LCOs was quite variable among samples but was in general higher in mycorrhized root exudates than in germinating spore exudates. The proposed structures are represented in Fig. 3.

Purification of AM fungal LCOs resulted in extremely low yields (picogram quantities). To facilitate the study of their biological activities, milligram quantities of Myc-LCO molecules were synthesized by bacterial genetic engineering (Supplementary Note 3), either by using appropriate *nod* gene mutants of *Sinorhizobium meliloti* and *Rhizobium leguminosarum*[24–26] (LCOs termed (Rhi)Myc-LCOs in the following text) or by the 'E. coli cell factory' procedure[27–29] (LCOs called (Syn)Myc-LCOs). For biological assays the two types of sulphated compound (with C16:0 or C18:1 N-acyl chains) were mixed (1/1) as sulphated LCOs, and the same was done with the non-sulphated LCOs.

Myc–LCOs stimulate AM formation

AM fungal LCOs, produced by these procedures, were used to study their influence on formation of mycorrhiza. *M. truncatula* seedlings were grown in test tubes on gellified slants[15] in which a 1/1 mixture of
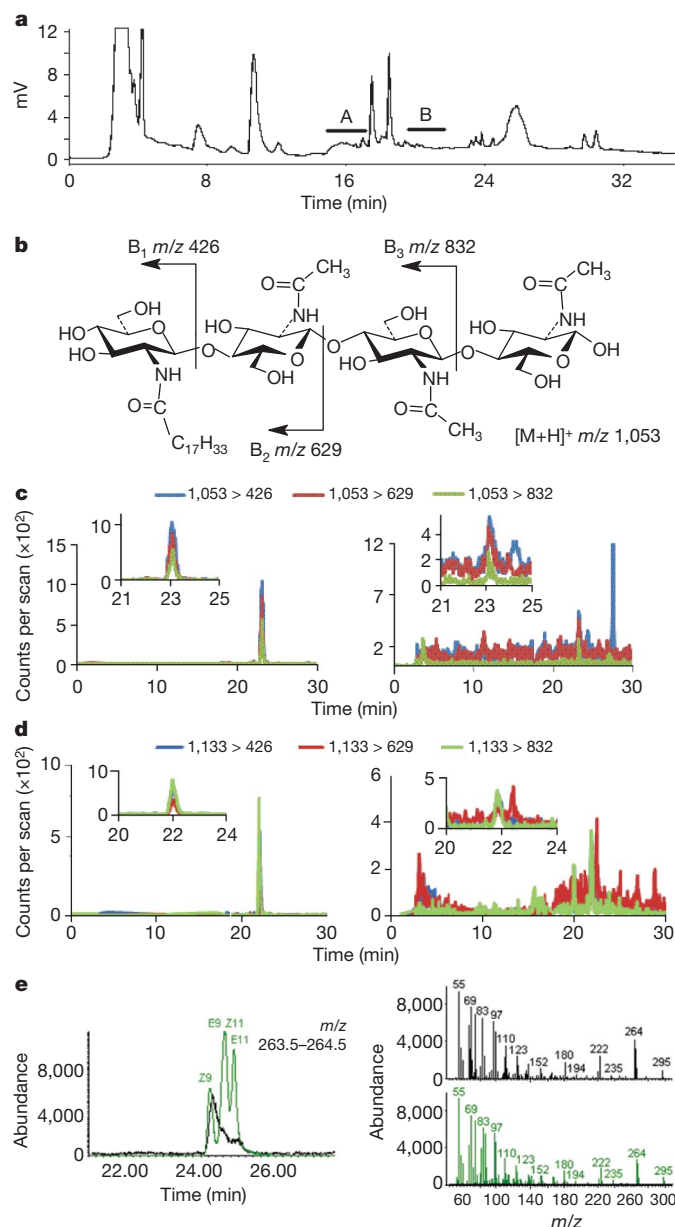
Figure 2 | Detection and characterization of LCOs in germinating *G. intraradices* spore exudates. a, Reverse-phase (C18) HPLC of butanol extract of exudates from 3 million germinating spores (GSE), showing two active fractions (A and B). **b**, LCO-IV(C18:1) molecule and B ions observed in LC/Q-Trap MS analysis in the positive mode. **c, d**, LC-QTRAP analysis in the MRM mode. **c**, Non-sulphated LCO-IV(C18:1) (transitions: blue, 1,053 > 426; red, 1,053 > 629; green, 1,053 > 832). Left, synthetic standard; right, GSE fraction B. **d**, Sulphated LCO-IV(C18:1) (transitions: blue, 1,133 > 426; red, 1,133 > 629; green, 1,133 > 832). Left, synthetic standard; right, GSE fraction A. The relevant part of the chromatograms is enlarged in the insets. **e**, Characterization of the fatty acyl C18:1 N-substitution (C18:1Δ9Z = oleic acid) of fraction B by positive electron impact mass spectrometry/gas chromatography. Left, chromatogram of fraction B hydrolysate (black) and four different standard C18:1 fatty acids (green). Right, electron impact spectrum of the fraction B hydrolysate, characteristic of oleic acid (top) and the corresponding standard (bottom).

Figure 3 | Chemical structures of natural and synthetic Myc-LCOs. a, b, Proposed chemical structures of two major Myc-LCOs, (**a**) LCO-IV(C16:0, S) and (**b**) LCO-IV(C18:1Δ9Z). **c**, General Myc-LCO structure. For both natural and synthetic Myc-LCOs $n = 1$ or 2, R2 = H or SO_3H. For natural Myc-LCOs, R1 = C16, C16:1, C16:2, C18:0 or C18:1Δ9Z (oleic acid); for synthetic Myc-LCOs, R1 = C16:0 or C18:1Δ9Z (oleic acid); and for rhizobial Myc-LCO analogues, R1 = C16:0 or C18:1 Δ11Z (*cis*-vaccenic acid).

interaction. Transcriptomic analysis of roots of model legumes has identified hundreds of genes upregulated by AM fungal infection or fungal diffusible factors[17,30]. Among these genes, ten were selected thanks to an *M. truncatula* transcriptome analysis using (Syn)Myc-LCOs (J. Becker and A. Niebel, personal communication). Quantitative PCR with reverse transcription showed that among the ten genes tested, four were upregulated in a *DMI3*-dependent manner by the (Syn)Myc-LCO treatment (Supplementary Note 4 and Supplementary Fig. 9), providing additional evidence that these molecules act as signals during the AM fungal symbiosis.

We then tested Myc-LCOs on the non-legume species *Tagetes patula* (Asteraceae) and *Daucus carota* (Umbelliferae). For *T. patula*, mycorrhization assays were performed on charred clay granules. Plantlets treated with a mixture (1/1) of sulphated and non-sulphated (Syn)Myc-LCOs showed a highly significant increase in the number of infection units per plant (Fig. 4b). Then, inoculated plants were treated with either pure sulphated or non-sulphated (Syn)Myc-LCOs, or with a 1/1 mixture of both. Treatment with the mixture elicited a doubling of the root colonization level (+104%), whereas pure non-sulphated and pure sulphated Myc-LCOs resulted in 75% and 42% increases, respectively (Fig. 4b).

sulphated and non-sulphated (Syn)Myc-LCOs (10 nM) was incorporated, and they were inoculated with sterile spores of *G. intraradices*. The number of infection units per plant (separate zones containing arbuscules and internal hyphal networks; Fig. 4a), as well as the density of infection sites per centimetre of root length, were greatly increased by the Myc-LCO treatment (Fig. 4a). We checked whether Myc-LCOs could induce *M. truncatula* genes involved at early steps of the

Figure 4 | Effect of Myc-LCOs on mycorrhization by *G. intraradices*. a, *M. truncatula*. Left, AM infection unit. Right, number of infection units per plant and density of infection in seedlings treated by sulphated and non-sulphated mixture of (Syn)Myc-LCOs (6 weeks, *n* = 15). **b**, *T. patula*. Left, effect of a sulphated and non-sulphated mixture of (Syn)Myc-LCOs on the number of infection units per plant (4 weeks, *n* = 12); right, percentage root colonization of plants treated by non-sulphated (NS), sulphated (S) and a 1/1 non-sulphated plus sulphated (NS + S) (Syn)Myc-LCO mixture (4 weeks, *n* = 40). **c**, Carrot excised transformed roots. Left, percentage colonization of roots treated by a non-sulphated plus sulphated mixture of (Rhi)Myc-LCO rhizobial analogues (8 weeks, *n* = 10). Right, percentage colonization of roots treated by a non-sulphated plus sulphated mixture of (Syn)Myc-LCOs (8 weeks, *n* = 15). In all experiments Myc-LCOs were used at 10 nM. Data are mean ± s.e.m. Statistics: **a**, Wilcoxon test; **b** (left) and **c**, *t*-test; **b** (right), Kruskal–Wallis; **0.001 < *P* < 0.01; *0.01 < *P* < 0.05.

AM fungi can form mycorrhiza on transformed root organ cultures[31]. To test whether Myc-LCOs could stimulate mycorrhization of such excised carrot roots, we used a 1/1 mixture of sulphated and non-sulphated (Rhi)Myc-LCO analogues, produced by rhizobial mutants (Supplementary Note 3 and Fig. 3). The incorporation of these compounds into the growth medium resulted in a very strong increase in colonization (+68%) (Fig. 4c). Similarly a significant increase of 20% was observed when a 1/1 mixture of synthetic sulphated and non-sulphated (Syn)Myc-LCOs was used. Therefore the response of roots to these Myc signals did not require the presence of aerial parts of the plant. On the three plant species, including legumes and non-legumes, synthetic Myc-LCOs enhance AM formation, probably by stimulating both root branching and infection density. This is strong evidence that the Myc-LCOs we have identified are genuine mycorrhizal signals.

Myc–LCOs stimulate root branching

AM fungi secrete diffusible compounds which stimulate root branching in *M. truncatula*, and this response can be genetically dissected[15]. We observed that HPLC fractions A and B, containing *G. intraradices* sulphated and non-sulphated LCOs respectively, elicited root-branching stimulation (RBS) (Supplementary Fig. 1b). To determine whether this was really due to LCOs and not to contaminating fungal compounds, we tested sulphated and non-sulphated (Syn)Myc-LCOs, separately or as a 1/1 mixture, at concentrations ranging from 10 nM to 10 pM on *M.*
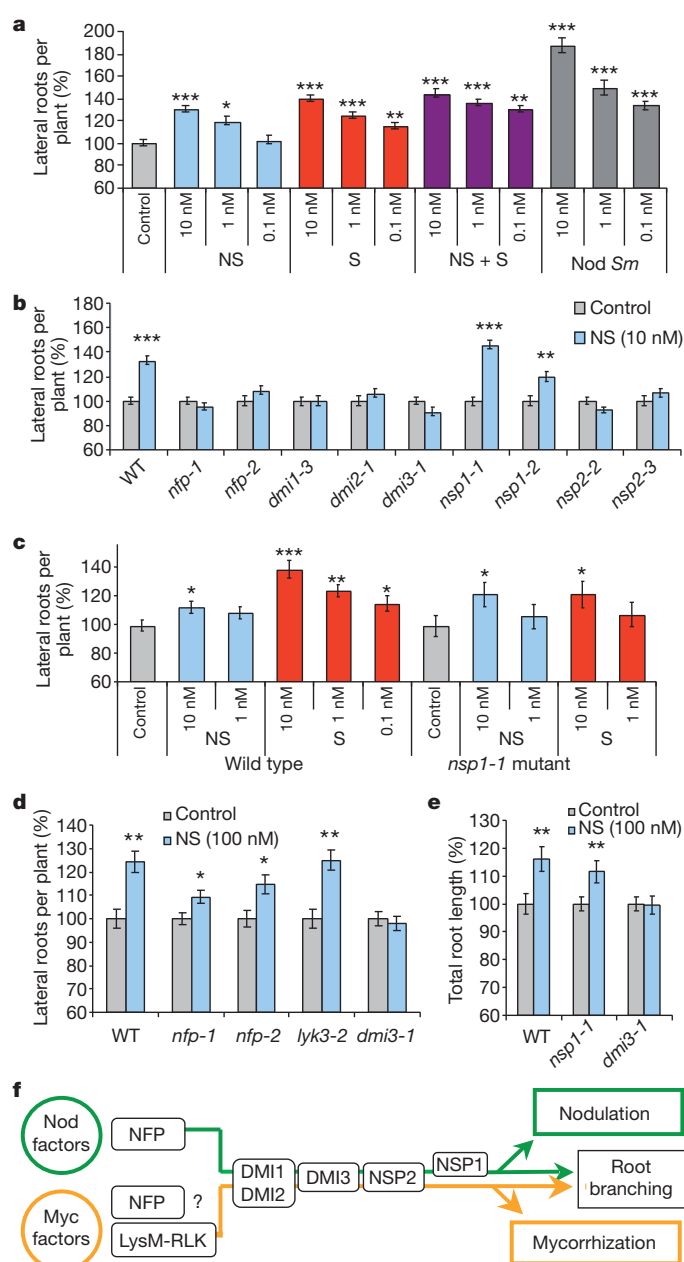
Figure 5 | Synthetic Myc-LCOs stimulate *M. truncatula* root branching by the DMI pathway. a, Effect of non-sulphated, sulphated, non-sulphated plus sulphated (Syn)Myc-LCOs and *S. meliloti* Nod factors (from 10 nM to 0.1 nM) on RBS in wild-type plants (*n* = 120). **b**, Effect of (Syn)Myc-LCOs in wild type and symbiotic signalling pathway mutants, with 10 nM NS (Syn)Myc-LCOs (*n* = 120). **c**, Effect of the *nsp1-1* mutation with non-sulphated or sulphated (Syn)Myc-LCOs (*n* = 80). **d**, Effect of 100 nM non-sulphated (Syn)Myc-LCOs in wild type and symbiotic mutants (*n* = 120). **e**, Effect of 100 nM non-sulphated (Syn)Myc-LCOs on total root length of wild type, *nsp1* and *dmi3* mutants (*n* = 80). **f**, Model for the genetic control of Nod- and Myc-factor-activated signal transduction pathways leading to RBS, nodulation and mycorrhization. Data are mean ± s.e.m. Statistics: **a**, **b**, analysis of variance; **c–e**, *t*-test; ****P* < 0.001; **0.001 < *P* < 0.01; *0.01 < *P* < 0.05. For root branching histograms, means of measures for days 5, 6, 7 and 8, are presented as percentages of the control value.

truncatula seedlings (Fig. 5a and Supplementary Fig. 10). Non-sulphated (Syn)Myc-LCOs elicited RBS at concentrations down to 10 nM/1 nM. Sulphated (Syn)Myc-LCOs and the mixture of the two types elicited RBS at concentrations down to 0.1 nM, and in some experiments down to 0.01 nM. Both sulphated and non-sulphated

Myc-LCOs were thus extremely active, but sulphated molecules were about 100-fold more active.

The symbiotic signalling pathway identified in *M. truncatula* includes genes coding for Nod factor perception (*NFP* and *LYK3*)[32,33], calcium signalling (*DMI1*, *DMI2* and *DMI3*)[34-36] and nodulation-specific transcription factors (*NSP1* and *NSP2*)[9,37,38]. All of these genes are required for nodulation whereas *DMI1*, *DMI2* and *DMI3* are also required for mycorrhization. They are all required, except *LYK3*, for RBS by *S. meliloti* Nod factors (ref. 15 and F. Maillet, personal communication). Sulphated Myc-LCOs, having structural similarities with *S. meliloti* Nod factors, might induce RBS by the Nod signalling pathway: to avoid possible cross talk between Nod and Myc pathways, we first tested non-sulphated (Syn)Myc-LCOs (at 10 nM) (Fig. 5b). The RBS response was dependent on the three *DMI* genes, which confirmed that Myc-LCOs are symbiotic signals. Downstream of *DMI3*, the response was independent of the *NSP1* gene, showing that non-sulphated Myc-LCOs do not trigger RBS by the Nod pathway. Surprisingly, the response was dependent on *NSP2* whereas *nsp2* mutants were previously reported to be Myc$^+$. We re-assessed the Myc phenotype by quantitatively measuring mycorrhiza formation at early stages: the *nsp2-2* mutant exhibited a highly significant 41% lower colonization level than wild-type plants, showing that *NSP2* is involved in Myc signalling (Supplementary Fig. 11). In addition, a mutation in *RAM1*, a gene specifically involved in Myc signalling, suppressed RBS (F. Maillet and G. Oldroyd, personal communication). All these data are consistent with non-sulphated Myc-LCOs stimulating root branching by a Myc pathway. Mutants of the two putative Nod factor receptor genes *LYK3* and *NFP*, acting upstream of *DMI* genes, are Myc$^+$. Mutation in *LYK3*, a gene involved in rhizobial recognition during infection, did not affect the RBS response to Myc-LCOs (Fig. 5d). The RBS response of *nfp* mutants varied with Myc-LCO concentration. At 10 nM, *nfp* mutants did not show a significant response (Fig. 5b). At 100 nM, whereas the *dmi3* mutant did not respond and the *lyk3* mutant exhibited a full RBS response as the wild type, *nfp* mutants showed a significant but intermediate stimulation (Fig. 5d).

With sulphated Myc-LCOs, the signalling leading to RBS seems to be by the Nod pathway (requirement for *NSP1*) (Fig. 5c and Supplementary Figs 12 and 13). The stimulating effect of sulphated Myc-LCOs at low concentrations, such as 0.1 nM (Fig. 5a and Supplementary Fig. 13), could be due to activation of the Nod pathway. Using the *nsp1-1* mutant to block Nod signalling, we observed that sulphated Myc-LCOs elicit RBS at concentrations similar to the non-sulphated ones, about 100-fold higher than the concentrations required to trigger RBS in wild-type plants (Fig. 5c). Thus, to elicit RBS by the Myc pathway, the required sulphated or non-sulphated Myc-LCO concentrations are about 100-fold higher than those required to elicit this response by the Nod pathway.

From an agronomical point of view, it was important to investigate the possible influence of Myc-LCOs on whole root development. To avoid a possible bias due to cross talk with the Nod pathway, non-sulphated compounds were used, as well as a mutant in the *NSP1* gene. At 100 nM, non-sulphated Myc-LCOs resulted in a highly significant 16% increase in total root length. This response was dependent on the *DMI3* gene but did not require the *NSP1* gene, specific to the Nod pathway (Fig. 5e).

Discussion

The discovery of the structure of diffusible Myc signals is an important milestone in our understanding of the evolution of plant root endosymbioses, and opens many possibilities for the molecular and cellular dissection of these interactions. The AM fungus *G. intraradices* secretes a mixture of sulphated and non-sulphated LCOs that have structural similarities with rhizobial Nod factors. In both Nod and Myc symbiotic signals, the chitin oligosaccharidic backbone is composed of four or five glucosamine residues, N-acylated on the glucosamine residue at the non-reducing end. *G. intraradices* Myc signals

have simpler structures than the Nod factors described so far[6,22] (Supplementary Note 1), and it is worth noting that rhizobial mutants that produce simple LCOs similar to Myc-LCOs are unable to penetrate into their legume hosts and to form nodules[24,25]. Interestingly, broad host-range rhizobia produce a mixture of sulphated and non-sulphated Nod factors[6,20,22].

The Myc signals that we have identified correspond to the definition of Myc factors that was proposed earlier[7-9,13]: (1) they elicit plant responses by the DMI pathway; (2) in a legume, downstream of DMI3 they activate a pathway that is NSP1 independent (a transcription activator specific to the nodulation pathway), and *RAM1* dependent (a gene specific to the Myc signalling pathway; F. Maillet and G. Oldroyd, personal communication). In addition they stimulate mycorrhiza formation in legumes and non-legumes. Other AM fungal signals inducing symbiotic gene expression or influencing root architecture have been reported, but the root responses did not require the *DMI* genes[39,40].

Genetic analysis of Myc signalling in a legume has revealed that the intricacy with Nod signalling is even more complex than previously thought (Fig. 5f). The NSP2 transcription activator now appears to be a component common to the Nod and Myc pathways. Mutants in the putative Nod factor receptor genes *NFP* and *LYK3*, not being altered for mycorrhiza formation, were not thought to be involved in Myc factor perception. This is the case for *LYK3*, but our data suggest that *NFP* is partly involved in the Myc-signal-elicited RBS response. Legumes have to discriminate the two types of symbiotic signal, despite their similarities, to activate the appropriate symbiotic programme. Nod factor and Myc factor receptors could discriminate their respective ligands by a combination of qualitative (structural) and quantitative differences (Supplementary Note 5).

The AM symbiosis is much more ancient than the rhizobium–legume association[18,19]. Bacterial *nod* genes required for LCO synthesis could have been acquired by lateral transfer from AM fungi about 60 million years ago and subsequently spread to a variety of soil bacteria[5,6,22]. Several data support the hypothesis that the Nod-factor-activated signalling pathway leading to legume nodulation is largely derived from a pre-existing signalling pathway leading to endomycorrhization[7,8,41-43]. Myc factors are not only symbiotic signals that stimulate mycorrhiza formation but also plant growth regulators. This suggests that the AM symbiosis, in the course of evolution, has selected fungal signals that can modify root development to facilitate further symbiotic infection.

The RBS assay, much simpler than the quantitative study of mycorrhization, could be used to facilitate genetic analysis of plant responsiveness to mycorrhizal signals. We have developed biotechnological procedures to produce Myc factors in large quantities. This opens the way to exploring the use of these very active natural molecules in agriculture.

METHODS SUMMARY

To detect fungal symbiotic signals, two bioassays were used: vetch root hair branching[20] and induction of a *pENOD11::GUS* construct in *M. truncatula* roots[21]. Three hundred litres of sterile exudates from mycorrhized carrot roots[44], and exudates from 40 million germinating spores of *G. intraradices*, were extracted with butanol, followed by solid-phase extraction on a C18 reverse-phase column eluted with different proportions of water/acetonitrile. The 50% acetonitrile fraction was purified by HPLC on a C18 reverse-phase column. Active fractions from mycorrhized root exudates were analysed by a UPLC-QToF mass spectrometer to determine accurate molecular masses. Exudates from germinating spores were analysed with the multiple-reaction monitoring device on a 4000 Q-Trap mass spectrometer. Milligram quantities of Myc-LCO analogues were synthesized by bacterial genetic engineering, either by using appropriate *nod* gene mutants of *S. meliloti* and *R. leguminosarum*[24-26] or by the 'E. coli cell factory' procedure[27-29]. Mycorrhization experiments were performed using sterile *G. intraradices* spores with seedlings of *M. truncatula*[15,45], *T. patula* and transformed root cultures of *D. carota* (carrot)[44]. Root colonization was estimated with a binocular microscope by counting the number of infection units per plant[15] or by the gridline intersect method[46]. The developmental activity of Myc-LCOs was studied using the *M.*

truncatula root branching bioassay[15]. Root system length was measured by image analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 29 October 2010.

1. Smith, S. E. & Read, D. J. in *Mycorrhizal Symbiosis*. 145–187 (Academic Press, 2008).
2. Harrison, M. J. Signaling in the arbuscular mycorrhizal symbiosis. *Annu. Rev. Microbiol.* **59,** 19–42 (2005).
3. Oldroyd, G. E. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59,** 519–546 (2008).
4. Stacey, G., Libault, M., Brechenmacher, L., Wan, J. & May, G. D. Genetics and functional genomics of legume nodulation. *Curr. Opin. Plant Biol.* **9,** 110–121 (2006).
5. Masson-Boivin, C., Giraud, E., Perret, X. & Batut, J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol.* **17,** 458–466 (2009).
6. Dénarié, J., Debellé, F. & Promé, J. C. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* **65,** 503–535 (1996).
7. Parniske, M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature Rev. Microbiol.* **6,** 763–775 (2008).
8. Oldroyd, G. E., Harrison, M. J. & Paszkowski, U. Reprogramming plant cells for endosymbiosis. *Science* **324,** 753–754 (2009).
9. Catoira, R. *et al.* Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell* **12,** 1647–1666 (2000).
10. Wais, R. J. *et al.* Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula. Proc. Natl Acad. Sci. USA* **97,** 13407–13412 (2000).
11. Kosuta, S. *et al.* Differential and chaotic calcium signatures in the symbiosis signaling pathway of legumes. *Proc. Natl Acad. Sci. USA* **105,** 9823–9828 (2008).
12. Oldroyd, G. E. & Downie, J. A. Nuclear calcium changes at the core of symbiosis signalling. *Curr. Opin. Plant Biol.* **9,** 351–357 (2006).
13. Riely, B. K., Ané, J. M., Penmetsa, R. V. & Cook, D. R. Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. *Curr. Opin. Plant Biol.* **7,** 408–413 (2004).
14. Weidmann, S. *et al.* Fungal elicitation of signal transduction-related plant genes precedes mycorrhiza establishment and requires the dmi3 gene in *Medicago truncatula. Mol. Plant Microbe Interact.* **17,** 1385–1393 (2004).
15. Oláh, B., Brière, C., Bécard, G., Dénarié, J. & Gough, C. Nod factors and a diffusible factor from arbuscular mycorrhizal fungi stimulate lateral root formation in *Medicago truncatula* via the DMI1/DMI2 signalling pathway. *Plant J.* **44,** 195–207 (2005).
16. Gutjahr, C. *et al.* Presymbiotic factors released by the arbuscular mycorrhizal fungus *Gigaspora margarita* induce starch accumulation in *Lotus japonicus* roots. *New Phytol.* **183,** 53–61 (2009).
17. Kuhn, H., Kuster, H. & Requena, N. Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in *Medicago truncatula. New Phytol.* **185,** 716–733 (2010).
18. Remy, W., Taylor, T. N., Hass, H. & Kerp, H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Natl Acad. Sci. USA* **91,** 11841–11843 (1994).
19. Sprent, J. I. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol.* **174,** 11–25 (2007).
20. Price, N. P. *et al.* Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are *O*-acetylated or sulphated. *Mol. Microbiol.* **6,** 3575–3584 (1992).
21. Andriankaja, A. *et al.* AP2-ERF transcription factors mediate Nod factor-dependent Mt *ENOD11* activation in root hairs via a novel *cis*-regulatory motif. *Plant Cell* **19,** 2866–2885 (2007).
22. D'Haeze, W. & Holsters, M. Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* **12,** 79R–105R (2002).
23. Roche, P., Lerouge, P., Ponthus, C. & Promé, J. C. Structural determination of bacterial nodulation factors involved in the *Rhizobium meliloti*-alfalfa symbiosis. *J. Biol. Chem.* **266,** 10933–10940 (1991).
24. Ardourel, M. *et al. Rhizobium meliloti* lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. *Plant Cell* **6,** 1357–1374 (1994).
25. Spaink, H. P. *et al.* Structural identification of metabolites produced by the NodB and NodC proteins of *Rhizobium leguminosarum. Mol. Microbiol.* **13,** 821–831 (1994).
26. Spaink, H. P., Wijfjes, A. H. & Lugtenberg, B. J. *Rhizobium* NodI and NodJ proteins play a role in the efficiency of secretion of lipochitin oligosaccharides. *J. Bacteriol.* **177,** 6276–6281 (1995).
27. Samain, E., Drouillard, S., Heyraud, A., Driguez, H. & Geremia, R. A. Gram-scale synthesis of recombinant chitooligosaccharides in *Escherichia coli. Carbohydr. Res.* **302,** 35–42 (1997).
28. Samain, E., Chazalet, V. & Geremia, R. A. Production of *O*-acetylated and sulfated chitooligosaccharides by recombinant *Escherichia coli* strains harboring different combinations of *nod* genes. *J. Biotechnol.* **72,** 33–47 (1999).
29. Ohsten Rasmussen, M., Hogg, B., Bono, J. J., Samain, E. & Driguez, H. New access to lipo-chitooligosaccharide nodulation factors. *Org. Biomol. Chem.* **2,** 1908–1910 (2004).
30. Liu, J. *et al.* Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. *Plant Cell* **15,** 2106–2123 (2003).
31. Bécard, G. & Fortin, J. A. Early events of vesicular–arbuscular mycorrhiza formation in Ri T-DNA transformed roots. *New Phytol.* **108,** 211–218 (1988).
32. Arrighi, J. F. *et al.* The *Medicago truncatula* lysin motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. *Plant Physiol.* **142,** 265–279 (2006).
33. Smit, P. *et al. Medicago* LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol.* **145,** 183–191 (2007).
34. Ané, J. M. *et al. Medicago truncatula* DMI1 required for bacterial and fungal symbioses in legumes. *Science* **303,** 1364–1367 (2004).
35. Endre, G. *et al.* A receptor kinase gene regulating symbiotic nodule development. *Nature* **417,** 962–966 (2002).
36. Lévy, J. *et al.* A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* **303,** 1361–1364 (2004).
37. Smit, P. *et al.* NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science* **308,** 1789–1791 (2005).
38. Kalo, P. *et al.* Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* **308,** 1786–1789 (2005).
39. Kosuta, S. *et al.* A diffusible factor from arbuscular mycorrhizal fungi induces symbiosis-specific *MtENOD11* expression in roots of *Medicago truncatula. Plant Physiol.* **131,** 952–962 (2003).
40. Gutjahr, C., Casieri, L. & Paszkowski, U. *Glomus intraradices* induces changes in root system architecture of rice independently of common symbiosis signaling. *New Phytol.* **182,** 829–837 (2009).
41. Chen, C., Gao, M., Liu, J. & Zhu, H. Fungal symbiosis in rice requires an ortholog of a legume common symbiosis gene encoding a Ca^{2+}/calmodulin-dependent protein kinase. *Plant Physiol.* **145,** 1619–1628 (2007).
42. Chen, C., Ané, J. M. & Zhu, H. OsIPD3, an ortholog of the *Medicago truncatula* DMI3 interacting protein IPD3, is required for mycorrhizal symbiosis in rice. *New Phytol.* **180,** 311–315 (2008).
43. Wang, B. *et al.* Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytol.* **186,** 514–525 (2010).
44. Chabot, S., Bécard, G. & Piche, Y. The life cycle of *Glomus intraradices* in root organ culture. *Mycologia* **84,** 315–321 (1992).
45. Ben Amor, B. *et al.* The *NFP* locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J.* **34,** 495–506 (2003).
46. Giovannetti, M. & Mosse, B. An evaluation of techniques for measuring vesicular-arbuscular mycorrhizal infection in roots. *New Phytol.* **84,** 489–500 (1980).

Author Contributions F.M. coordinated experiments for all bioassays, designed and performed the VsHab assay, and designed and performed *M. truncatula* bioassays with O.A. O.A. and V.P. purified fungal Myc-LCOs. M.G. and A.H. prepared germinating spores. L.C. and A.H. performed the *ENOD11* bioassay. A.H. designed and performed mycorrhization tests on carrots and *Tagetes*. O.A. performed statistical analyses. F.M. and D.G. extracted Myc-LCO analogues from rhizobial mutant cultures, and E.A.M. and H.D. synthesized Myc-LCOs by the cell factory technique. For mass spectrometry, UPLC/QToF was performed by V.P. and V.P.-P., and QTRAP by V.P.-P. Quantitative PCR with reverse transcription experiments were designed and analysed by A.N. and realized by D.F. and O.A. G.B. supervised devising, planning and interpreting experiments with AM fungal material (material production, spore germination and mycorrhization tests). J.D. conceived and directed the project, and wrote the manuscript with the help of F.M., V.P. and G.B.

METHODS

Exudates from mycorrhized roots and from germinating spores. Exudates of carrot roots mycorrhized by *G. intraradices* were purchased from Premier Tech Biotechnologies. Purified sterile spores of *G. intraradices* were purchased from Premier Tech Biotechnologies and from Agronutrition. Spores were germinated at 30 °C in a 2% CO_2 incubator for 10 days.

Plant material. Seeds of *M. truncatula* var. Jemalong A17 wild type[9] and symbiotically defective mutants derived from this line[9,32], *nfp-1* (C31), *nfp-2*, *lyk3-2* (W1), *dmi1-3* (Y6), *dmi2-1* (TR25), *dmi3-1* (TRV25) and *nsp1-1* (B85) were provided by C. Gough. Seeds of the *nsp2-2* and *nsp2-3* mutants[47] were provided by G. Oldroyd. Seeds of *V. sativa* subsp. *nigra* were multiplied by SCJ INRA Dijon. Seeds of the *M. truncatula* line carrying the *pENOD11::GUS* fusion[48] were provided by E. Journet. Seeds of *T. patula* (French marigold) var. Légion d'honneur were from Caillard. Sterile cultures of carrot hairy root were obtained as previously described[44].

Bioassays for purification of AM fungal symbiotic signals. The root hair branching assay of vetch[20] was used with the following modifications. Young seedlings were left 3 days in Petri dishes, on Fahraeus agar, in a vertical position in a growth chamber (22 °C). Observations of root hairs were performed in a 0.02% methylene blue solution under a light microscope, 30 h after treatment (40 µl per seedling). The *MtENOD11* gene induction assay[21] was used with no paper on the agar surface and 40-µl treatment per seedling root.

Myc-LCO purification. Active LCO compounds were extracted as previously described[49] with the following modifications. No ammonium acetate was added to the eluents and a pre-purification step was used with a solid-phase extraction on a Discovery DSC C18 5-µm (Supelco) column, with three elution steps: 20%, 50% and 100% ACN in water. Samples were collected every minute along the HPLC gradient.

UPLC/Q-ToF MS analyses. The UPLC column was a C18 Acquity (2.1 mm × 100 mm, 1.7 µm, Waters) and the flow rate was 0.45 ml min^{-1}. For the more hydrophilic compounds (semi-preparative HPLC fractions 1–9), the programme was a linear gradient ranging from 10% ACN in 1% acetic acid/water to 100% ACN within 7 min, followed by an isocratic step at 100% ACN for 2 min, 2 min return to the initial conditions and 1 min stabilization. The more hydrophobic compounds (HPLC fractions 6–11) were separated using a linear gradient from 25% ACN in 0.1% acetic acid/water to 100% ACN within 7 min, followed by similar reconditioning. Ten microlitres were injected.

The mass spectrometer was a Q-ToF Premier (Waters). The capillary was set to 3.2 kV and the cone to 10 V. Internal lock mass was performed by continuous introduction of leucine-enkephalin. Analyses were performed in the negative and positive modes. For fragmentation experiments, a collision energy of 15 V was used.

LC/Q-Trap analyses in the MRM mode. The binary HPLC pump (Agilent) was equipped with a C18 Acclaim 120 column (2.1 mm × 250 mm, 5 µm, Dionex). Separation started at 50% ACN in water for 5 min, followed by a 25-min gradient to 100% ACN maintained for 5 min, at a constant flow rate of 200 µl min^{-1}. Ten-microlitre samples were injected. The mass spectrometer was a 4000 Q-Trap (Applied Biosystems) with a Turbo V (MDS Sciex) source. Samples were monitored in the positive MRM mode as reported previously[50]. The capillary voltage was set to 4,000 V. Collision energy was between 25 and 65 V and declustering potential was between 60 and 75 V, optimized for each molecule. The MRM channels were set according to the transitions of molecular ion $[M + H]^+$ to the fragment ions corresponding to the loss of one, two or three GlcNAc at the reducing end (sulphated or not): LCO-IV(C16:0), 1,027 > 806, 1,027 > 603, 1,027 > 400; LCO-IV(C16:0, S), 1,107 > 806, 1,107 > 603, 1,107 > 400; LCO-IV(C18:1), 1,053 > 832, 1,053 > 629, 1,053 > 426; LCO-IV(C18:1, S), 1,133 > 832, 1,133 > 629, 1,133 > 426.

Biochemical characterization of Myc-LCOs. To remove the sulphate group, mild hydrolysis was performed as previously described[23]. For the fatty-acid analysis of Myc-LCOs, HPLC fractions (F8–F11) of butanol extracts of germinating spore (33 million spores) exudates, were dissolved in 4 M KOH and heated at 100 °C for 3 h, neutralized and extracted with chloroform. Methylation of the fatty acids was done by addition of diazomethane. Analysis was on a GC/MS system (5973N, Hewlett Packard). GC: He = 1 ml min^{-1}; programme, 70 °C (3 min), up to 100 °C at 5 °C min^{-1}, then to 240 °C at 3 °C min^{-1} and finally at 10 °C min^{-1} to 300 °C maintained for 3 min. MS: positive electron impact 70 eV; source, 230 °C; quadripole, 150 °C.

Production of synthetic Myc-LCOs and Myc-LCO analogues. Rhizobial *nod* mutants and the 'E. coli cell factory' method were used to produce milligram quantities of Myc-LCO analogues. Sulphated Myc-LCO analogues were purified from liquid cultures of the *S. meliloti nodFL* mutant (strain GMI 6629 containing

the pMH682 plasmid), as previously reported[24,51]. For the production of non-sulphated Myc-LCO analogues, the *R. leguminosarum* strain LPR5045 (pMP247) was used[25]. LCOs were purified as previously reported[25]. Synthesis of Myc-LCOs by the 'E. coli cell factory' was as described in the literature[28,29] using the *S. meliloti nodBC* and *nodH* genes. LCO purity control and structural determination were performed by mass spectrometry (UPLC/Q-ToF, LC/Q-Trap) and NMR[52] spectroscopy (Supplementary Fig. 14). Their biological activity was checked with the *ENOD11::GUS* and vetch root hair deformation assays (Supplementary Fig. 15).

Mycorrhization. The effect of exogenous LCO addition on AM formation was tested in non-optimal conditions, to avoid the possible risk that the fungal endogenous synthesis of LCOs might be sufficient to ensure an optimal AM formation. In these conditions, mycorrhization percentages were expected to be low. For mycorrhization of *M. truncatula* in test-tubes, 50 sterile spores were put at the bottom of each slope of medium containing Myc-LCOs or control solution. For mycorrhization of *T. patula* on charred clay granules (Oil Dry US Special, Brenntag) hydrated with Long Ashton low phosphate solution[53], 50-ml Falcon tubes, pierced with three small holes at the bottom, were filled with sterilized substrate and individually placed in 120-ml plastic boxes (5.5 cm diameter, 7 cm height), closed with an opaque pierced cap, filled with 80 ml water and wrapped in aluminium foil. For inoculation of *T. patula*, spores were suspended in 1 ml of 100 nM Myc-LCOs (to obtain a 10 nM final concentration) or control solution and 100 spores were dropped around the seed. Each plant received 1 ml of treatment solution, twice a week for 3 weeks. Pots were placed in a growth chamber (25 °C, 16 h photoperiod, 180 µE m^{-2} s^{-1}) for 4 weeks. For mycorrhization of excised carrot roots, root fragments were laid on Petri dishes (diameter = 90 mm) containing a first layer of 20 ml M medium (0.3% Phytagel) and a second layer of the same medium containing 20 or 200 spores per millilitre and Myc-LCOs at the appropriate concentration. Dishes were incubated in the dark (24 °C, 50% humidity). Myc-LCOs were added once a week on the plate surface during the first 3 or 4 weeks. Root colonization was estimated under a binocular microscope, after staining root fragments with Schaeffer black ink[54], by (1) counting infection units (discrete zones containing arbuscules) per plant[15] or (2) estimating the percentage of root length colonized by the fungus by the gridline intersection method[46].

Root branching assay. This was performed with *M. truncatula* seedlings as previously reported[15] but vitamins were not added to the M medium and plates were incubated at 20 °C. For measurement of root length we used an Epson scanner 10000 XL and the image analysis system of Winrhizo Scientific Software (Instruments Regent).

Statistics. *M. truncatula* mycorrhization data were analysed by the non-parametric Wilcoxon test. Statistical software was from the R system[55]. *M. truncatula* root length, carrot root and *T. patula* mycorrhization data were analysed by Student's *t*-test. For root branching data, statistical analyses were performed with Statgraphics Centurion (version XV) software (Statpoint Technologies). Data were tested by analysis of variance with a general linear model procedure of data from several (mostly three) independent experiments. To analyse repeated measure experiments, the statistical general linear model was built to take into account the impact of four factors on variability: experiments, treatments, plants and days (5–8). The normality of residuals was verified by the Kolmogorov–Smirnov test.

47. Hirsch, S. *et al.* GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell* **21,** 545–557 (2009).
48. Journet, E. P. *et al. Medicago truncatula* ENOD11: a novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells. *Mol. Plant Microbe Interact.* **14,** 737–748 (2001).
49. Poinsot, V. *et al.* Unusual methyl-branched α,β-unsaturated acyl chain substitutions in the Nod factors of an arctic rhizobium, *Mesorhizobium* sp. strain N33 (*Oxytropis arctobia*). *J. Bacteriol.* **183,** 3721–3728 (2001).
50. Gomez-Roldan, V. *et al.* Strigolactone inhibition of shoot branching. *Nature* **455,** 189–194 (2008).
51. Demont, N., Debellé, F., Aurelle, H., Dénarié, J. & Promé, J. C. Role of the *Rhizobium meliloti* nodF and nodE genes in the biosynthesis of lipo-oligosaccharidic nodulation factors. *J. Biol. Chem.* **268,** 20134–20142 (1993).
52. Olsthoorn, M. M. *et al.* Novel branched nod factor structure results from alpha-(1→3) fucosyl transferase activity: the major lipo-chitin oligosaccharides from Mesorhizobium loti strain NZP2213 bear an alpha-(1→3) fucosyl substituent on a nonterminal backbone residue. *Biochemistry* **37,** 9024–9032 (1998).
53. Hewitt, E. J. *Sand and Water Culture Methods Used in the Study of Plant Nutrition* (Commonwealth Agricultural Bureaux, 1966).
54. Vierheilig, H., Coughlan, A. P., Wyss, U. & Piche, Y. Ink and vinegar, a simple staining technique for arbuscular-mycorrhizal fungi. *Appl. Environ. Microbiol.* **64,** 5004–5007 (1998).
55. R Development Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2009).

LETTER

A high C/O ratio and weak thermal inversion in the atmosphere of exoplanet WASP-12b

Nikku Madhusudhan[1]†, Joseph Harrington[2], Kevin B. Stevenson[2], Sarah Nymeyer[2], Christopher J. Campo[2], Peter J. Wheatley[3], Drake Deming[4], Jasmina Blecic[2], Ryan A. Hardy[2], Nate B. Lust[2], David R. Anderson[5], Andrew Collier-Cameron[6], Christopher B. T. Britt[2], William C. Bowman[2], Leslie Hebb[7], Coel Hellier[5], Pierre F. L. Maxted[5], Don Pollacco[8] & Richard G. West[9]

The carbon-to-oxygen ratio (C/O) in a planet provides critical information about its primordial origins and subsequent evolution. A primordial C/O greater than 0.8 causes a carbide-dominated interior, as opposed to the silicate-dominated composition found on Earth[1]; the atmosphere can also differ from those in the Solar System[1,2]. The solar C/O is 0.54 (ref. 3). Here we report an analysis of dayside multi-wavelength photometry[4,5] of the transiting hot-Jupiter WASP-12b (ref. 6) that reveals C/O ≥ 1 in its atmosphere. The atmosphere is abundant in CO. It is depleted in water vapour and enhanced in methane, each by more than two orders of magnitude compared to a solar-abundance chemical-equilibrium model at the expected temperatures. We also find that the extremely irradiated atmosphere (T > 2,500 K) of WASP-12b lacks a prominent thermal inversion (or stratosphere) and has very efficient day–night energy circulation. The absence of a strong thermal inversion is in stark contrast to theoretical predictions for the most highly irradiated hot-Jupiter atmospheres[7–9].

The transiting hot Jupiter WASP-12b orbits a star slightly hotter than the Sun (6,300 K) in a circular orbit at a distance of only 0.023 astronomical units (AU), making it one of the hottest exoplanets known[6]. Thermal emission from the dayside atmosphere of WASP-12b has been reported using the Spitzer Space Telescope[10], at 3.6 μm, 4.5 μm, 5.8 μm and 8 μm wavelengths[4], and from ground-based observations in the J (1.2 μm), H (1.6 μm) and Ks (2.1 μm) bands[5] (Fig. 1).

The observations provide constraints on the dayside atmospheric composition and thermal structure, based on the dominant opacity source in each bandpass. The J, H and Ks channels[5] have limited molecular absorption features, and hence probe the deep layers of the planetary atmosphere, at pressure $P \approx 1$ bar, where the temperature $T \approx 3,000$ K (Fig. 1). The Spitzer observations[4], on the other hand, are excellent probes of molecular composition. CH_4 has strong absorption features in the 3.6-μm and 8-μm channels, CO has strong absorption in the 4.5-μm channel, and H_2O has its strongest feature in the 5.8-μm channel and weaker features in the 3.6-μm, 4.5-μm and 8-μm channels. The low brightness temperatures in the 3.6-μm (2,700 K) and 4.5-μm (2,500 K) channels, therefore, clearly suggest strong absorption due to CH_4 and CO, respectively. The high brightness temperature in the 5.8-μm channel, on the other hand, indicates low absorption due to H_2O. The strong CO absorption in the 4.5-μm channel also indicates temperature decreasing with altitude, because a thermal inversion would cause emission features of CO in the same channel with a significantly higher flux than at 3.6 μm (refs 11 and 12).

The broadband observations allow us to infer the chemical composition and temperature structure of the dayside atmosphere of WASP-12b using a statistical retrieval technique[11]. We combined a one-dimensional atmosphere model with a Markov-chain Monte Carlo sampler[11,13] that computes over 4×10^6 models to explore the

parameter space. The phase space included thermal profiles with and without inversions, and equilibrium and non-equilibrium chemistry over a wide range of atomic abundances. Our models include the dominant sources of infrared opacity in the temperature regime of WASP-12b (refs 14, 15 and 16): H_2O, CO, CH_4, CO_2, H_2–H_2 collision-induced absorption, and TiO and VO where the temperatures
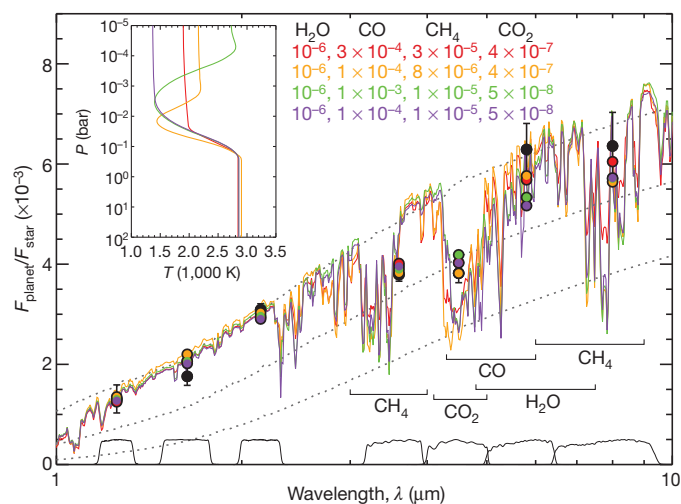
Figure 1 | Observations and model spectra for dayside thermal emission of WASP-12b. *F* is the flux. The black filled circles with error bars show the data with 1 s.d. errors: four Spitzer observations[4] (3.6 μm, 4.5 μm, 5.8 μm and 8 μm), and three ground-based observations in the J (1.2 μm), H (1.6 μm) and Ks (2.1 μm) bands[5]. Four models fitting the observations are shown in the coloured solid curves in the main panel, and the coloured circles are the channel-integrated model points. The corresponding temperature profiles are shown in the inset. The molecular compositions are shown as number ratio with respect to molecular hydrogen; all the models have C/O between 1 and 1.1. The thin grey dotted lines show the blackbody spectra of WASP-12b at 2,000 K (bottom), 2,500 K (middle) and 3,000 K (top). A Kurucz model[29] was used for the stellar spectrum, assuming uniform illumination over the planetary disk (that is, weighted by 0.5; ref. 7). The black solid lines at the bottom show the photometric band-passes in arbitrary units. The low fluxes at 3.6 μm and 4.5 μm are explained by methane and CO absorption, respectively, required for all the models that fit. The high flux in the 5.8-μm channel indicates less absorption due to H_2O. The observations can be explained to high precision by models without thermal inversions. Models with strong thermal inversions are ruled out by the data (see Fig. 3). The green model features a thermal inversion at low pressures ($P < 0.01$ bar), but the corresponding spectrum is almost indistinguishable from the purple model, which does not have a thermal inversion; both models have identical compositions and identical thermal profiles for $P > 0.01$ bar. Thus, any potential thermal inversion is too weak to be detectable by current instruments.

[1]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Planetary Sciences Group, Department of Physics, University of Central Florida, Orlando, Florida 32816-2385, USA. [3]Department of Physics, University of Warwick, Coventry, CV4 7AL, UK. [4]NASA's Goddard Space Flight Center, Greenbelt, Maryland 20771-0001, USA. [5]Astrophysics Group, Keele University, Staffordshire ST5 5BG, UK. [6]School of Physics and Astronomy, University of St Andrews, North Haugh, Fife KY16 9SS, UK. [7]Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA. [8]Astrophysics Research Centre, School of Mathematics and Physics, Queen's University, University Road, Belfast, BT7 1NN, UK. [9]Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK. †Present address: Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA.

are high enough for them to exist in the gas phase[7,17]. The host star has a significantly enhanced metallicity ($2 \times$ solar)[6], and evolutionary processes can further enhance the abundances[18,19]; Jupiter has $3 \times$ solar C/H (ref. 18). Our models therefore explore wide abundance ranges: 0.01–$100 \times$ solar for C/H and O/H, and 0.1–$10 \times$ solar for C/O. Figure 2 shows the mixing ratios of H_2O, CO, CH_4 and CO_2 and the ratios of C/H, O/H and C/O required by the models at different levels of fit. Figure 3 presents the temperature profiles.

We find a surprising lack of water and overabundance of methane (Fig. 2). At 2,000–3,000 K, assuming solar abundances yields CO and H_2O as the dominant species besides H_2 and He (refs 15 and 16). Most of the carbon, and the same amount of oxygen, are present in CO, and some carbon exists as CH_4. The remaining oxygen in a hydrogen-dominated atmosphere is mostly in H_2O; small amounts are also present in species such as CO_2. The CO/H_2 and H_2O/H_2 mixing ratios should each be $>5 \times 10^{-4}$, CH_4/H_2 should be $<10^{-8}$, and CO_2/H_2 should be about 10^{-8}, under equilibrium conditions at a nominal pressure of 0.1 bar. The requirement of $H_2O/H_2 \leq 6 \times 10^{-6}$ and $CH_4/H_2 \geq 8 \times 10^{-6}$ (both at 3σ, 99.73% significance; Fig. 2) is therefore inconsistent with equilibrium chemistry using solar abundances.

The observations place a strict constraint on the C/O ratio. We detect C/O ≥ 1 at 3σ significance (Fig. 2). Our results rule out a solar C/O of 0.54 at 4.2σ. Our calculations of equilibrium chemistry[16,20] using C/O $= 1$ yield mixing ratios of H_2O, CO and CH_4 that are consistent with the observed constraints. We find that, for C/O $= 1$, H_2O mixing ratios as low as 10^{-7} and CH_4 mixing ratios as high as 10^{-5} can be attained at the 0.1–1 bar level for temperatures around 2,000 K and higher. And, although the CO mixing ratio is predicted to be $>10^{-4}$, making it the dominant molecule after H_2 and He, CO_2 is predicted to be negligible ($<10^{-9}$). The theoretical predictions for a

C/O $= 1$ atmosphere are consistent with the observed constraints on H_2O, CH_4, CO and CO_2 (Fig. 2).

The observations rule out a strong thermal inversion deeper than 0.01 bar (Fig. 3). Thermal inversions at lower pressures have opacities too low to induce features in the emission spectrum that current instruments can resolve. For comparison[11,12,21], all stratospheric inversions in Solar System giant planets, and those consistent with hot-Jupiter observations, exist at pressures between 0.01 bar and 1 bar. The major contributions to all the observations come from the lower layers of the atmosphere, $P > 0.01$ bar, where we rule out a thermal inversion (Supplementary Fig. 1). The observations also suggest very efficient day–night energy redistribution (Fig. 2). The low brightness temperatures at 3.6 μm and 4.5 μm imply that only part of the incident stellar energy is re-radiated from the dayside, whereas up to 45% is absorbed and redistributed to the nightside. The possibility of a deep thermal inversion and inefficient redistribution was suggested recently[5], based on observations in the J, H and Ks channels, but the Spitzer observations rule out both conditions.

The lack of a prominent thermal inversion contrasts with recent work that designates WASP-12b as a member of the class of very hot Jupiters that are expected to host inversions[7,22]. At $T > 2,000$ K, molecules such as TiO and VO, which are strong absorbers in the ultra-violet/visible, are expected to be in gas phase and potentially cause thermal inversions[7]. WASP-12b, now being the hottest planet without a distinct inversion, presents a major challenge to existing atmospheric classification schemes for exoplanets based on thermal inversions[7,22]. Although there are hints of low chromospheric activity[8] in the host star, it remains to be seen whether the high incident continuum ultraviolet flux expected for WASP-12b might be efficient in photo-dissociating inversion-causing compounds, thus explaining the lack of a strong
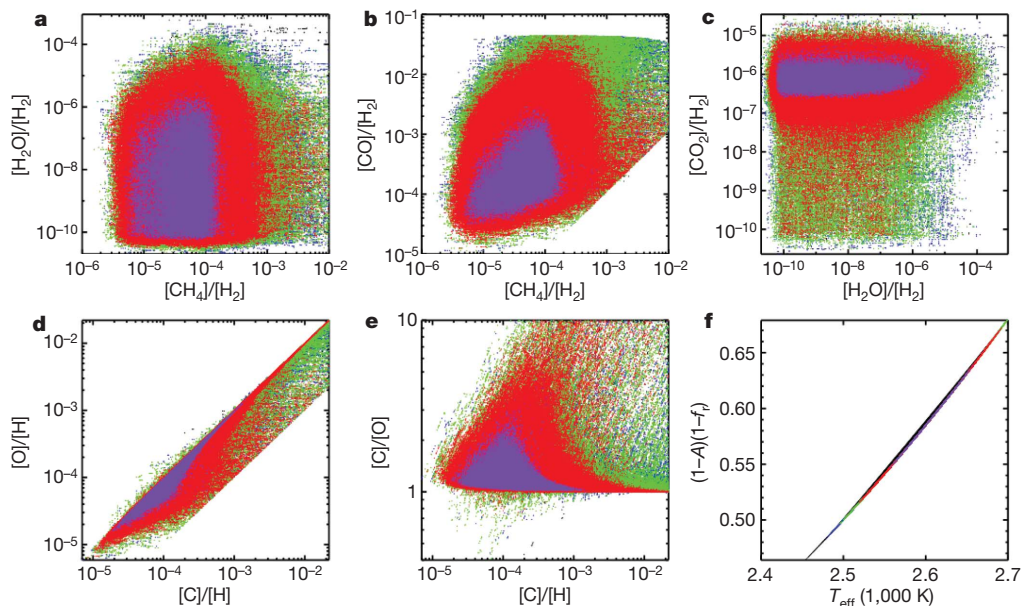
Figure 2 | Constraints on the atmospheric composition of WASP-12b. a–e The distributions of models fitting the seven observations (Fig. 1) at different levels of χ^2 are shown. The coloured dots show χ^2 surfaces, with each dot representing a model realization. The purple, red, green, blue and black colours correspond to models with χ^2 less than 7, 14, 21 and 28 and $\chi^2 > 28$, respectively (χ^2 ranges between 4.8–51.3). Mixing ratios are shown as ratios by number with respect to H_2. At 3σ significance, the constraints on the composition are $H_2O/H_2 \leq 6 \times 10^{-6}$, $CH_4/H_2 \geq 8 \times 10^{-6}$, CO/ $H_2 \geq 6 \times 10^{-5}$, $CO_2/H_2 \leq 5 \times 10^{-6}$, and C/O > 1. The compositions of the best-fitting models (with $\chi^2 < 7$) span $H_2O/H_2 = 5 \times 10^{-11}$ to 6×10^{-6}, CO/ $H_2 = 3 \times 10^{-5}$ to 3×10^{-3}, $CH_4/H_2 = 4 \times 10^{-6}$ to 8×10^{-4} and $CO_2/$ $H_2 = 2 \times 10^{-7}$ to 7×10^{-6}; the corresponding ranges in C/O and elemental abundances are C/O $= 1$ to 6.6, C/H $= 2 \times 10^{-5}$ to 10^{-3} and O/H $= 2 \times 10^{-5}$

to 10^{-3}. The constraints on the C/H and O/H ratios are governed primarily by the constraints on CO, which is the dominant molecule after H_2 and He. On the basis of thermochemical equilibrium, the inferred CH_4/H_2 and H_2O/H_2 mixing ratios are possible only for C/O ≥ 1, consistent with our detection of C/O ≥ 1. f, Constraints on the day–night energy redistribution[11], given by $(1 - A)(1 - f_r)$, where A is the bond albedo and f_r is the fraction of incident energy redistributed to the nightside. Up to $f_r = 0.45$ is possible (for $A = 0$). Thus, the observations support very efficient redistribution. An additional observation in the z' (0.9 μm) band was reported recently[30]. However, the observation implies a value for the orbital eccentricity inconsistent with other data in the literature[4,5]. We therefore decided to exclude this observation from the analysis presented here, although including it does not affect our conclusions regarding the value of C/O or the temperature structure.
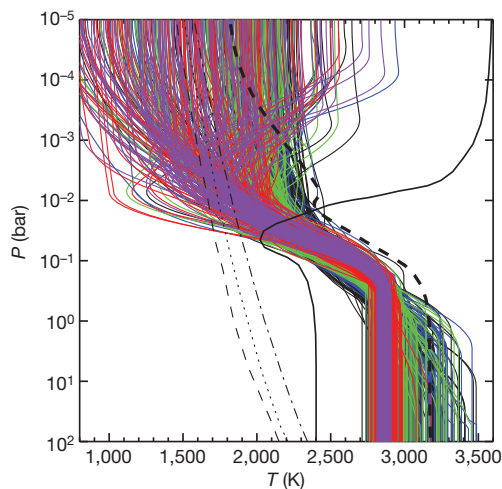
Figure 3 | Thermal profiles of WASP-12b. The solid thin lines show profiles at different degrees of fit (description of colours is as in Fig. 2); only 100 randomly chosen profiles for each χ^2 level are shown, for clarity. The thick black solid curve in the front shows a published profile from a self-consistent model of WASP-12b with a thermal inversion, adapted from ref. 17, which assumes solar abundances. The thick black dashed curve shows the same model but without a thermal inversion. If a thermal inversion is present in WASP-12b, it is expected to be prominent, as shown by the thick solid black curve. A prominent thermal inversion between 0.01 bar and 1 bar is ruled out by the data at 4σ. The ostensibly large inversions in the figure are at low pressures (below 0.01 bar), which have low optical depths, and hence minimal influence on the emergent spectrum (see Fig. 1). The observations are completely consistent with thermal profiles having no inversions. Small thermal inversions are also admissible by the data, and could potentially result from dynamics. The thin black lines show the condensation curves of TiO at solar (dotted), $0.1 \times$ solar (dashed) and $10 \times$ solar (dash-dotted) compositions[17].

inversion[8]. Alternatively, the amount of vertical mixing might be insufficient to keep TiO/VO aloft in the atmosphere to cause thermal inversions[17]. A C/O = 1 might also yield lower TiO/VO than that required to cause a thermal inversion. It is unlikely that the TiO/VO in WASP-12b might be lost to cold traps[17], given the high temperatures in the deep atmosphere on the dayside and nightside.

If high C/O ratios are common, then the formation processes and compositions of extrasolar planets are probably very different from expectations based on Solar System planets. The host star has super-solar metallicity but initial analyses find its C/O consistent with solar[6,23]. In the core accretion model, favoured for the formation of Jupiter, icy planetesimals containing heavy elements coalesce to form the core, followed by gas accretion[19,24]. The abundances of elemental oxygen and carbon are enhanced equally[18,19], maintaining a C/O like the star's. If the host star had a C/O ≈ 1, then the C/O we detect in WASP-12b would have been evident. However, if the stellar C/O is indeed <1, then the C/O enhancement in WASP-12b's atmosphere would suggest either an unexpected origin for the planetesimals, a local overdensity of carbonaceous grains[2,25], or a different formation mechanism entirely. Although carbon-rich giant planets like WASP-12b have not been observed, theory predicts myriad compositions for carbon-dominated solid planets[1,2]. Terrestrial-sized carbon planets, for instance, could be dominated by graphite or diamond interiors, as opposed to the silicate composition of Earth[1,2]. If carbon dominates the heavy elements in the interior of a hot Jupiter, estimates of mass and radius could change compared to those based on solar abundances. Future interior models[26] should investigate the contribution of high C/O to the large radius of WASP-12b: that is, 1.75 Jupiter radii for 1.4 Jupiter masses (ref. 6).

The observed molecular abundances in the dayside atmosphere of WASP-12b motivate the exploration of a new regime in atmospheric chemistry. It remains to be seen whether photochemistry in WASP-12b

can significantly alter the composition in the lower layers of the atmosphere at P = 0.01–1 bar; these layers contribute most to the observed spectrum (Supplementary Fig. 1). Explaining the observed composition as a result of photochemistry with solar abundances would be challenging. CH_4 is more readily photodissociated than H_2O (refs 9 and 27), and hence a depletion of CH_4 over that predicted with solar abundances might be expected, as opposed to the observed enhancement of CH_4. Apart from the spectroscopically dominant molecules considered in this work, minor species such as OH, C_2H_2 and FeH (refs 27 and 28), which are not detectable by current observations, could potentially be measured with high-resolution spectroscopy in the future. Detection of these species would allow additional constraints on equilibrium and non-equilibrium chemistry in WASP-12b, although their effect on the C/O would be negligible. Models of exoplanetary atmospheres have typically assumed solar abundances and/or solar C/O, thereby exploring a very limited region of parameter space[7,12,20]. Data sufficient for a meaningful constraint on C/O exist for only a few exoplanets. That this initial C/O statistical analysis has C/O ≥ 1 potentially indicates a wide diversity of planetary compositions.

1. Bond, J. C., O'Brien, D. P. & Lauretta, D. S. The compositional diversity of extrasolar planets. I. *In situ* simulations. *Astrophys. J.* **715**, 1050–1070 (2010).
2. Kuchner, M. & Seager, S. Extrasolar carbon planets. Preprint at ⟨http://arxiv.org/abs/astro-ph/0504214⟩ (2005).
3. Asplund, M., Grevesse, N. & Sauval, A. in. *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis* (eds Barnes, T. G. III & Bash, F. N.) 25–38 (ASP Conf. Ser. 336, 2005).
4. Campo, C. *et al.* On the orbit of exoplanet WASP-12b. *Astrophys. J.* (in the press). Preprint at ⟨http://arxiv.org/abs/1003.2763⟩ (2010).
5. Croll, B. *et al.* Near-infrared thermal emission from WASP-12b: detections of the secondary eclipse in Ks, H & J. *Astrophys. J.* (in the press). Preprint at ⟨http://arxiv.org/abs/1009.0071⟩ (2010).
6. Hebb, L. *et al.* 2009, WASP-12b: the hottest transiting extrasolar planet yet discovered. *Astrophys. J.* **693**, 1920–1928 (2009).
7. Fortney, J. J., Lodders, K., Marley, M. S. & Freedman, R. S. A unified theory for the atmospheres of the hot and very hot Jupiters: two classes of irradiated atmospheres. *Astrophys. J.* **678**, 1419–1435 (2008).
8. Knutson, H. A., Howard, A. W. & Isaacson, H. A correlation between stellar activity and hot Jupiter emission spectra. *Astrophys. J.* **720**, 1569–1576 (2010).
9. Zahnle, K. *et al.* Atmospheric sulfur photochemistry on hot Jupiters. *Astrophys. J.* **701**, L20–L24 (2009).
10. Werner, M. W. *et al.* The Spitzer Space Telescope mission. *Astrophys. J. Suppl. Ser.* **154**, 1–9 (2004).
11. Madhusudhan, N. & Seager, S. A temperature and abundance retrieval method for exoplanet atmospheres. *Astrophys. J.* **707**, 24–39 (2009).
12. Burrows, A., Budaj, J. & Hubeny, I. Theoretical spectra and light curves of close-in extrasolar giant planets and comparison with data. *Astrophys. J.* **678**, 1436–1457 (2008).
13. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (eds) *Markov Chain Monte Carlo in Practice* (Chapman & Hall, 1996).
14. Swain, M. R. *et al.* Molecular signatures in the near-infrared dayside spectrum of HD 189733b. *Astrophys. J.* **690**, L114–L117 (2009).
15. Lodders, K. & Fegley, B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. I. Carbon, nitrogen, and oxygen. *Icarus* **155**, 393–424 (2002).
16. Burrows, A. & Sharp, C. M. Chemical equilibrium abundances in brown dwarf and extrasolar giant planet atmospheres. *Astrophys. J.* **512**, 843–863 (1999).
17. Spiegel, D. S., Silverio, K. & Burrows, A. Can TiO explain thermal inversions in the upper atmospheres of irradiated giant planets? *Astrophys. J.* **699**, 1487–1500 (2009).
18. Atreya, S. K. & Wong, A. S. Coupled clouds and chemistry of the giant planets—a case for multiprobes. *Space Sci. Rev.* **116**, 121–136 (2005).
19. Owen, T. *et al.* A low-temperature origin for the planetesimals that formed Jupiter. *Nature* **402**, 269–270 (1999).
20. Seager, S. *et al.* On the dayside thermal emission of hot Jupiters. *Astrophys. J.* **632**, 1122–1131 (2005).
21. Yung, Y. & DeMore, W. B. *Photochemistry of Planetary Atmospheres* (Oxford University Press, 1999).
22. Hubeny, I., Burrows, A., & Sudarsky, D. A possible bifurcation in atmospheres of strongly irradiated stars and planets. *Astrophys. J.* **594**, 1011–1018 (2003).
23. Fossati, L. *et al.* A detailed spectropolarimetric analysis of the planet-hosting star WASP-12. *Astrophys. J.* **720**, 872–886 (2010).
24. Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
25. Lodders, K. Jupiter formed with more tar than ice. *Astrophys. J.* **611**, 587–597 (2004).

26. Fortney, J. J., Marley, M. S., &. Barnes, J. W. Planetary radii across five orders of magnitude in mass and stellar insolation: application to transits. *Astrophys. J.* **659,** 1661–1672 (2007).

27. Line, M. R., Liang, M. C. & Yung, Y. L. High-temperature photochemistry in the atmosphere of HD 189733b. *Astrophys. J.* **717,** 496–502 (2010).

28. Cushing, M. C., Rayner, J. T. & Vacca, W. D. An infrared spectroscopic sequence of M, L, and T dwarfs. *Astrophys. J.* **623,** 1115–1140 (2005).

29. Castelli, F. & Kurucz, R. L. New grids of ATLAS9 model atmospheres. Preprint at ⟨http://arxiv.org/abs/astro-ph/0405087⟩ (2004).

30. Lopez-Morales, M. *et al.* Day-side z′-band emission and eccentricity of WASP-12b. *Astrophys. J.* **716,** L36–L40 (2010).

LETTER

Evidence for a spin–aligned neutron–proton paired phase from the level structure of ^{92}Pd

B. Cederwall[1], F. Ghazi Moradi[1], T. Bäck[1], A. Johnson[1], J. Blomqvist[1], E. Clément[2], G. de France[2], R. Wadsworth[3], K. Andgren[1], K. Lagergren[1,4], A. Dijon[2], G. Jaworski[5,6], R. Liotta[1], C. Qi[1], B. M. Nyakó[7], J. Nyberg[8], M. Palacz[5], H. Al-Azri[3], A. Algora[9], G. de Angelis[10], A. Ataç[11], S. Bhattacharyya[2]†, T. Brock[3], J. R. Brown[3], P. Davies[3], A. Di Nitto[12], Zs. Dombrádi[7], A. Gadea[9], J. Gál[7], B. Hadinia[1], F. Johnston-Theasby[3], P. Joshi[3], K. Juhász[13], R. Julin[14], A. Jungclaus[15], G. Kalinka[7], S. O. Kara[11], A. Khaplanov[1], J. Kownacki[5], G. La Rana[12], S. M. Lenzi[16], J. Molnár[7], R. Moro[12], D. R. Napoli[10], B. S. Nara Singh[3], A. Persson[1], F. Recchia[16], M. Sandzelius[1]†, J.-N. Scheurer[17], G. Sletten[18], D. Sohler[7], P.-A. Söderström[8], M. J. Taylor[3], J. Timár[7], J. J. Valiente-Dobón[10], E. Vardaci[12] & S. Williams[19]

Shell structure and magic numbers in atomic nuclei were generally explained by pioneering work[1] that introduced a strong spin–orbit interaction to the nuclear shell model potential. However, knowledge of nuclear forces and the mechanisms governing the structure of nuclei, in particular far from stability, is still incomplete. In nuclei with equal neutron and proton numbers ($N = Z$), enhanced correlations arise between neutrons and protons (two distinct types of fermions) that occupy orbitals with the same quantum numbers. Such correlations have been predicted to favour an unusual type of nuclear superfluidity, termed isoscalar neutron–proton pairing[2–6], in addition to normal isovector pairing. Despite many experimental efforts, these predictions have not been confirmed. Here we report the experimental observation of excited states in the $N = Z = 46$ nucleus ^{92}Pd. Gamma rays emitted following the ^{58}Ni(^{36}Ar,2n)^{92}Pd fusion–evaporation reaction were identified using a combination of state-of-the-art high-resolution γ-ray, charged-particle and neutron detector systems. Our results reveal evidence for a spin-aligned, isoscalar neutron–proton coupling scheme, different from the previous prediction[2–6]. We suggest that this coupling scheme replaces normal superfluidity (characterized by seniority coupling[7,8]) in the ground and low-lying excited states of the heaviest $N = Z$ nuclei. Such strong, isoscalar neutron–proton correlations would have a considerable impact on the nuclear level structure and possibly influence the dynamics of rapid proton capture in stellar nucleosynthesis.

For all known nuclei, including those residing along the $N = Z$ line up to around mass 80, a detailed analysis of properties such as binding energies[9] and the spectroscopy of excited states[10] strongly suggests that normal isovector (isospin $T = 1$, see Fig. 1) pairing is dominant at low excitation energies. On the other hand, there are long-standing predictions[2–6] for a change in the heavier $N = Z$ nuclei, from a nuclear superfluid dominated by isovector pairing to a structure where isoscalar ($T = 0$) neutron–proton (np) pairing has a major influence, as the mass number increases towards the exotic doubly magic nucleus ^{100}Sn, the heaviest $N = Z$ nucleus predicted to be bound.

Nuclei with $N = Z$ and mass number >90 can only be produced in the laboratory with very low cross-sections. The related problems of identifying and distinguishing such reaction products and their associated γ-rays from the vast array of $N > Z$ nuclei that are present in much greater numbers from the reactions used have prevented observation of their low-lying excited states until now. In the present work, the experimental difficulties have been overcome through the use of a highly efficient detector system and a prolonged experimental running period.

Excited states in ^{92}Pd were populated following heavy-ion fusion-evaporation reactions at GANIL (Grand Accélérateur National d'Ions Lourds), France. ^{36}Ar ions, accelerated to a kinetic energy of 111 MeV, were used to bombard an isotopically enriched (99.83%) ^{58}Ni target. Light charged particles (mainly protons and α-particles), neutrons and γ-rays emitted in the reactions were detected in coincidence. A schematic layout of the experimental set-up is shown in Fig. 2.

The two-neutron (2n) evaporation reaction channel following formation of the ^{94}Pd compound nucleus, leading to ^{92}Pd, was very weakly populated, with a relative yield of less than 10^{-5} of the total fusion cross-section. Gamma-rays from decays of excited states in ^{92}Pd were identified by comparing γ-ray spectra in coincidence with two emitted neutrons and no charged particles with γ-ray spectra in coincidence with other combinations of neutrons and charged particles. The typical efficiency for detecting any charged particle was 66%. This number rises to 88% or higher if more than one such particle is emitted in a particular reaction channel. The clean identification of neutrons is crucial, as scattering of neutrons from one detector segment to another can be misinterpreted as two neutrons; this would give rise to a background from the much more prolific reaction channels (where only one neutron has been emitted) in γ-ray spectra gated by two neutrons. But because neutrons have a finite velocity, the difference in detection
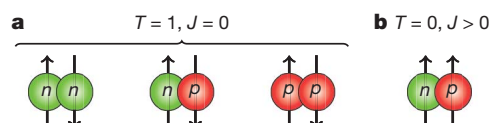
Figure 1 | Schematic illustration of the two possible pairing schemes in nuclei. a, The normal isospin $T = 1$ triplet. The two like-particle pairing components are responsible for most known effects of nuclear superfluidity. Within a given nucleus these isovector components are restricted to spin zero owing to the Pauli principle. **b,** Isoscalar $T = 0$ neutron–proton pairing. Here the Pauli principle allows only non-zero components of angular momentum.

[1]Department of Physics, Royal Institute of Technology, SE-10691 Stockholm, Sweden. [2]Grand Accélérateur National d'Ions Lourds (GANIL), CEA/DSM – CNRS/IN2P3, F-14076 Caen Cedex 5, France. [3]Department of Physics, University of York, York YO10 5DD, UK. [4]Joint Institute for Heavy-Ion Research, Holifield Radioactive Ion Beam Facility, Oak Ridge, Tennessee 37831, USA. [5]Heavy Ion Laboratory, University of Warsaw, 02-093 Warsaw, Poland. [6]Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland. [7]Institute of Nuclear Research of the Hungarian Academy of Sciences, ATOMKI, H-4001 Debrecen, Hungary. [8]Department of Physics and Astronomy, Uppsala University, SE-75121 Uppsala, Sweden. [9]IFIC, CSIC, University of Valencia, E-46071 Valencia, Spain. [10]Instituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Legnaro, I-35020 Legnaro, Italy. [11]Department of Physics, Ankara University, 06100 Tandogan Ankara, Turkey. [12]Dipartimento di Scienze Fisiche, Università di Napoli and Instituto Nazionale di Fisica Nucleare, I-80126 Napoli, Italy. [13]Department of Information Technology, University of Debrecen, H-4010 Debrecen, Hungary. [14]Department of Physics, University of Jyväskylä, FIN-40014 Jyväskylä, Finland. [15]Instituto de Estructura de la Materia, CSIC, E-28006 Madrid, Spain. [16]Dipartimento di Fisica dell'Università di Padova and Instituto Nazionale di Fisica Nucleare, Sezione di Padova, I-35122 Padova, Italy. [17]Université Bordeaux 1, CNRS/IN2P3, Centre d'Etudes Nucléaires de Bordeaux Gradignan, F-33175 Gradignan, France. [18]The Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark. [19]TRIUMF, Vancouver, British Columbia V6T 2A3, Canada. †Present addresses: VECC, 1/AF Bidhan Nagar, Kolkata 700064, India (S.B.); Department of Physics, University of Jyväskylä, FIN-40014 Jyväskylä, Finland (M.S.).
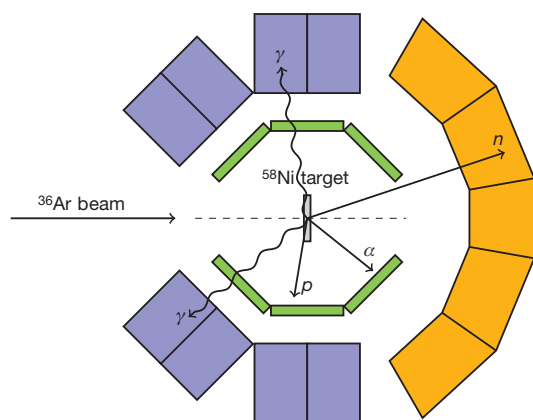
Figure 2 | Schematic illustration of the experimental set-up used to identify γ-ray transitions from excited states in ^{92}Pd. The light particles and γ-rays emitted from the ^{36}Ar + ^{58}Ni reaction were observed using three different detector systems. The innermost detector array, DIAMANT[16,17] (green), which consisted of 80 CsI scintillators, was used to detect light charged particles, mainly α-particles and protons, and acted as a veto detector in the selection of events with no charged particles emitted. The Neutron Wall[18] (orange), comprising 50 liquid scintillator detectors and covering a solid angle of 1π in the forward direction, was used for the detection of evaporated neutrons. It is able to discriminate between neutron and γ-ray interactions by means of a combined time-of-flight and pulse-shape analysis technique. Gamma-rays emitted from the reaction products were detected using the EXOGAM[19,20] high-purity Ge detector system (blue). Seven segmented Ge clover detectors were placed at an angle of 90° and four detectors at an angle of 135° relative to the beam direction, leaving room for the Neutron Wall at forward angles.

time is typically smaller for interactions resulting from two separate neutrons compared to a single scattered neutron. Background contributions from neutron scattering in 2*n*-gated spectra were significantly reduced by applying a criterion to the time difference in the time-of-flight parameter, relative to the distance between the neutron detectors firing. After such corrections, the efficiency for correctly identifying both neutrons from a 2*n*-event was 3%. Figure 3a–c shows projected γ-ray spectra from the charged particle-vetoed, 2*n*-selected E_γ–E_γ coincidence matrix when γ-rays coincident with the 874 keV, 912 keV and 750 keV transitions (Fig. 3a–c, respectively) assigned to ^{92}Pd are selected. By comparing spectra with and without the charged particle veto condition applied, it is clear that these γ-rays are not associated with emission of charged particles from the compound nucleus. Figure 3d shows a plot of the intensity ratios of the 874 keV, 912 keV and 750 keV γ-rays (filled circles) in coincidence with two neutrons and one neutron, respectively, proving that the γ-rays assigned to ^{92}Pd belong to the 2*n*-evaporation reaction channel. An extensive literature search was also performed in order to exclude the possibility that the γ-rays assigned to ^{92}Pd could be due to the decay of excited states in some other nucleus. In particular, γ-rays from reactions involving possible target impurities were taken into account. See Supplementary Information for further details on the data analysis.

The three most intense γ-ray transitions assigned to ^{92}Pd (874 keV, 912 keV and 750 keV) have been ordered into a ground-state band based on their relative intensities (Fig. 3e). The uncertainties in the relative intensities of the γ-ray transitions translate into a corresponding uncertainty in their ordering, and consequently, also in the absolute positions of the 2^+ and 4^+ states. As shown in Fig. 3, these γ-rays form a mutually coincident decay sequence. Although the limited statistics precludes an accurate angular distribution analysis and hence firm spin assignments, it is likely that the 874 keV, 912 keV and 750 keV γ-ray transitions constitute a cascade of stretched E2 transitions depopulating the first excited 2^+, 4^+ and 6^+ states, respectively (Fig. 3e).

Nuclei immediately below ^{100}Sn on the Segré (*N*,*Z*) chart, with *N*, $Z < 50$, may show special structural features, as the valence neutrons and protons here can move in identical orbits. Here, for the heaviest
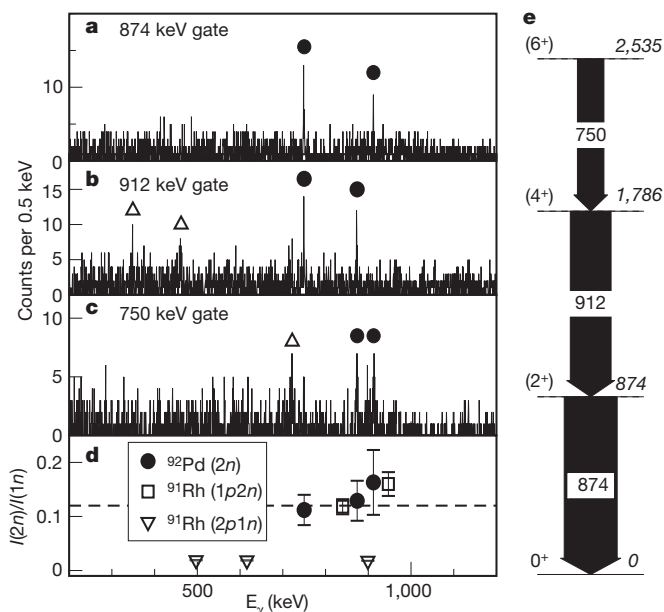

Figure 3 | Identification of γ-ray transitions in ^{92}Pd. a–c, Gamma-ray energy spectra detected in coincidence with the 874 keV, 912 keV and 750 keV γ-rays, with the additional requirement that two neutrons and no charged particle(s) were detected in coincidence. These γ-rays, assigned to depopulate the 2^+, 4^+ and 6^+ states in ^{92}Pd, respectively, are marked by filled circles. Gamma-rays from the ^{36}Ar-induced 1*p*1*n*-evaporation reaction on small amounts of carbon deposited on the target during irradiation, leading to the production of ^{46}V nuclei, are visible in **b** and **c** (open triangles). These γ-rays appear in the projected spectra owing to a combined effect of the limited detection efficiency for charged particles, the finite neutron/γ separation in the neutron detectors, the presence of γ-ray transitions at 914.9 keV and 750.7 keV in the level scheme of ^{46}V (refs 21, 22), and the fact that the reaction products from carbon contamination may recoil out of the target material, leading to Doppler broadening of such γ-rays. **d**, Intensity ratios of the γ-rays assigned to ^{92}Pd in coincidence with two neutrons and one neutron, respectively. The dashed line indicates the value expected for γ-rays in coincidence with two neutrons, obtained from the relative 1*n*- and 2*n*-detection efficiencies. Measured intensity ratios for γ-rays from previously known reaction products (^{91}Ru[23] and ^{91}Rh[24]) are included for comparison. **e**, Level scheme assigned to ^{92}Pd. The assigned spin-parity (left) and level energy (right, in keV) is given for each level. The energies (in keV) and relative intensities (in %, normalized to the intensity of the 874 keV transition) of the γ-ray transitions assigned to ^{92}Pd are as follows: 873.6(2), 100(8); 912.4(2), 77(5); 749.8(3), 50(6). Given uncertainties are standard (statistical) errors.

$N \approx Z$ nuclei, state-of-the-art shell model calculations predict the appearance of ground-state and low-lying yrast structures based on spin-aligned systems of np pairs, similar to a scenario proposed more than four decades ago[11]. The np-paired ground-state configuration emanates from the strong attractive interaction between $g_{9/2}$ neutrons and protons in aligned angular momentum ($J = 9$) coupling, and is hence different from the predictions of a BCS type of isoscalar np pairing condensate in $N \approx Z$ nuclei[2–6]. The shell model calculations were performed using empirical two-body matrix elements in the $f_{5/2}$ $p_{3/2}$ $p_{1/2}$ $g_{9/2}$ model space; see Supplementary Information for details.

In Fig. 4c we show the results, compared with experimental data for ^{92}Pd (this work), ^{94}Pd (ref. 12) and ^{96}Pd (ref. 13). The level structure of the semi-magic ($N = 50$) nucleus ^{96}Pd, with four proton holes relative to the $Z = 50$ closed shell core, exhibits the typical traits of a nucleus in the normal isovector pairing phase for which the seniority coupling scheme dominates. A transition from the ground state to the first excited 2^+ state requires the breaking of one $g_{9/2}$ proton-hole pair, and therefore the energy spacing between these two levels is rather large. The distance between the subsequent levels gradually decreases as the angular momentum vectors of the $g_{9/2}$ quasiproton holes align
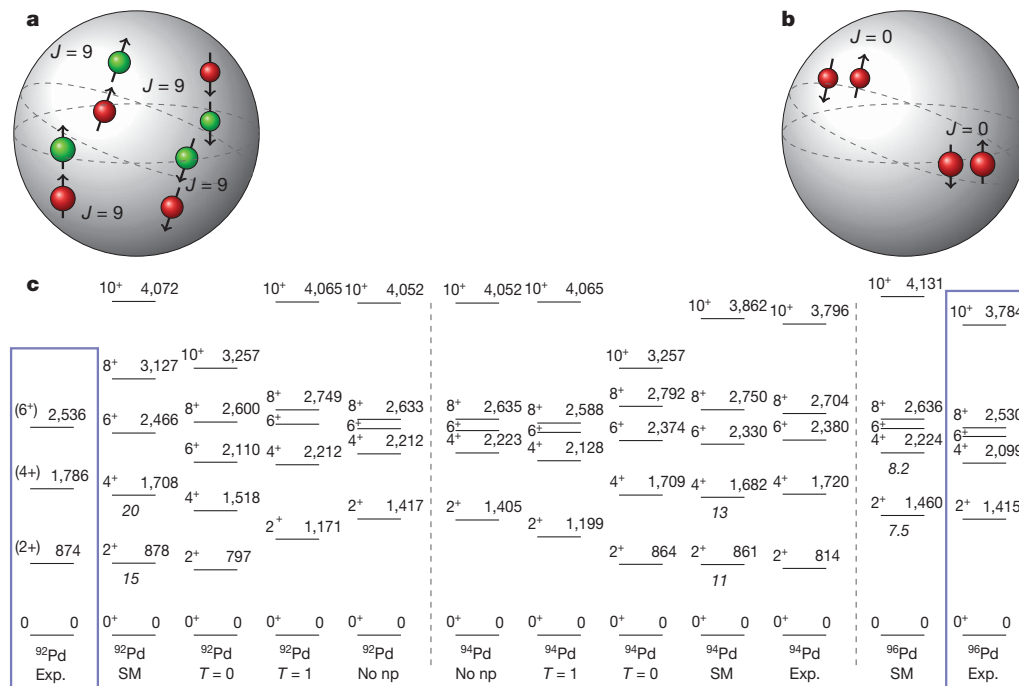
Figure 4 | Illustration of the predicted ground-state wavefunctions of ^{92}Pd and ^{96}Pd, and comparison of calculated and experimental level energies in ^{92}Pd, ^{94}Pd and ^{96}Pd. a, Schematic illustration of the structure of the ground-state wavefunction of ^{92}Pd in the spin-aligned np paired phase (green, neutron hole; red, proton hole). The main component of the wavefunction can be viewed as a system of deuteron-like np hole pairs with respect to the $^{100}_{50}$Sn$_{50}$ 'core', spinning around the centre of the nucleus. **b**, As **a** but for ^{96}Pd in the normal pairing phase. **c**, Experimental level energies (keV) in the ground-state bands of ^{92}Pd (present work) and ^{94,96}Pd (refs 12, 13) compared with shell model predictions. Calculated $B(E2: 2^+ \rightarrow 0^+)$ and $B(E2: 4^+ \rightarrow 2^+)$ values

(Weisskopf units) are also shown in italics below the corresponding initial levels. Text at bottom of each set of levels shows nuclide and gives further details. The calculated spectra for ^{92,94}Pd include, in addition to full neutron–proton interactions (SM), also results for pure $T = 0$ and pure $T = 1$ neutron–proton interactions. The results obtained without residual neutron-proton interactions (no np: that is, normal seniority coupling involving only isovector, $T = 1$, nn and pp pairing), are also shown for ^{92,94}Pd. The vertical dashed lines separate information for different nuclides. Blue boxes highlight the experimental data for ^{92}Pd and ^{96}Pd, emphasizing their different structures.

until the 8^+ state is reached. Here, the angular momentum vectors of one pair of proton holes are maximally aligned, and in order to reach higher-lying states the other proton-hole pair has to be broken. This spin sequence terminates in the 12^+ state, where all four proton holes in the $g_{9/2}$ orbital are fully aligned. In contrast, the calculated spectrum of ^{92}Pd, with four proton holes and four neutron holes relative to the ^{100}Sn core, has a nearly constant energy spacing between consecutive levels. To examine the influence from different components of the np interaction on this spectrum, we also performed the same calculation while including only the $T = 0$ component of the interaction matrix elements, including only the $T = 1$ components, or excluding all np interactions (that is, all $T_z = 0$ components). As seen in Fig. 4c, the calculated spectrum of ^{92}Pd for the latter case has a strong resemblance to the spectrum of the closed neutron shell nucleus ^{96}Pd. For the full calculation (case SM in Fig. 4c), the calculated energy spectra agree very well with those deduced from experiment. It is evident that the $T = 0$ component of the np interactions plays a dominating role for the spectrum of ^{92}Pd, whereas such interactions between the valence nucleons are absent in ^{96}Pd. The calculated wavefunctions for the ground state and low-lying yrast states in ^{92}Pd are completely dominated by the isoscalar np pairs in the spin-aligned $J^\pi = 9^+$ coupling. The nucleus ^{94}Pd represents an interesting intermediate case.

A simple semiclassical picture of the isoscalar spin-aligned coupling scheme can help to illustrate why the low-lying yrast states in ^{92}Pd are nearly equidistant. It is a consequence of the variation in the spatial overlap between the valence particle wavefunctions as the angular momentum vectors of np hole pairs circling in one direction align with the angular momentum vectors of those circling in the opposite direction, yielding the total angular momentum $2\hbar$, $4\hbar$, $6\hbar$, and so on. This variation is approximately linear for small values of angular

momentum. This mechanism for generating the total angular momentum in the nucleus is quite different from those present in normal superfluid nuclei. The regularly spaced level sequence observed in the full calculation for ^{92}Pd is therefore a distinct signature of the spin-aligned isoscalar mode, in the absence of collective vibrational excitations (see Supplementary Information for further details). The fact that the ordering of the experimentally observed γ-ray transitions is affected by some uncertainty does not change the interpretation of the data. The effect of a different ordering would be a maximal change in the 2^+ and 4^+ energies by 124 keV and 162 keV, respectively, still in approximate agreement with the theoretical prediction. The special topology of the ground-state wavefunction predicted for ^{92}Pd is illustrated schematically in Fig. 4a, and may be compared with the case for the normal pairing phase in ^{96}Pd (Fig. 4b). In the spin-aligned np paired phase, the main component of the nuclear ground-state wavefunction can, in a semiclassical picture, be regarded as built of a system of deuteron-like np hole pairs spinning around the core, each with maximum angular momentum. The special character of the wavefunction also implies a deformed intrinsic structure.

Although the experimental data presented in this Letter strongly suggest that a spin-aligned neutron–proton paired phase is present in ^{92}Pd, further experimental information is needed to confirm this interpretation. In particular, measurements of particle transfer reactions, extraction of electromagnetic transition rates ($B(E2; 0^+ \rightarrow 2^+)$ values) using Coulomb excitation, and precise mass measurements would help to elucidate the structural evolution of nuclei along the $N = Z$ line and to develop a better understanding of neutron–proton correlations and their implications for nuclear shell structure far from stability. This is also of importance for understanding the reaction rates as well as the end point of the astrophysical rapid proton capture process[14,15], which

may influence the composition and X-ray burst profiles of accreting neutron stars, and the nucleosynthesis of neutron deficient isotopes.

1. Goeppert Mayer, M. On closed shells in nuclei. II. *Phys. Rev.* **75,** 1969–1970 (1949).
2. Engel, J., Langanke, K. & Vogel, P. Pairing and isospin symmetry in proton-rich nuclei. *Phys. Lett. B* **389,** 211–216 (1996).
3. Goodman, A. L. Restoration of axial symmetry of the equilibrium shape of ^{24}Mg by pairing correlations. *Adv. Nucl. Phys.* **11,** 260–263 (1979).
4. Engel, J. *et al.* Neutron-proton correlations in an exactly solvable model. *Phys. Rev. C* **55,** 1781–1788 (1997).
5. Satula, W. & Wyss, R. Competition between T = 0 and T = 1 pairing in proton-rich nuclei. *Phys. Lett. B* **393,** 1–6 (1997).
6. Civitarese, O., Reboiro, M. & Vogel, P. Neutron-proton pairing in the BCS approach. *Phys. Rev. C* **56,** 1840–1843 (1997).
7. Talmi, I. Generalized seniority and structure of semi-magic nuclei. *Nucl. Phys. A* **172,** 1–24 (1971).
8. Talmi, I. Generalized seniority states with definite isospin. *Nucl. Phys. A* **686,** 217–240 (2001).
9. Macchiavelli, A. O. *et al.* Is there np pairing in N=Z nuclei? *Phys. Rev. C* **61,** 041303(R) (2000).
10. Afanasjev, A. & Frauendorf, S. Description of rotating N=Z nuclei in terms of isovector pairing. *Phys. Rev. C* **71,** 064318 (2005).
11. Danos, M. & Gillet, V. Stretch scheme, a shell model description of deformed nuclei. *Phys. Rev.* **161,** 1034–1044 (1967).
12. Mărginean, N. *et al.* Yrast isomers in ^{95}Ag, ^{95}Pd, and ^{94}Pd. *Phys. Rev. C* **67,** 061301 (2003).
13. Alber, D., Bertschat, H. H., Grawe, H., Haas, H. & Spellmeyer, B. Nuclear structure studies of the neutron deficient N=50 nucleus ^{96}Pd. *Z. Phys. A* **332,** 129–135 (1989).
14. Schatz, H. *et al.* End point of the rp process on accreting neutron stars. *Phys. Rev. Lett.* **86,** 3471–3474 (2001).
15. Clement, R. R. C. *et al.* New approach for measuring properties of rp-process nuclei. *Phys. Rev. Lett.* **92,** 172502 (2004).
16. Scheurer, J. N. *et al.* Improvements in the in-beam γ-ray spectroscopy provided by an ancillary detector coupled to a Ge -spectrometer: the DIAMANT-EUROGAM II example. *Nucl. Instrum. Methods Phys. Res. A* **385,** 501–510 (1997).
17. Gál, J. *et al.* The VXI electronics of the DIAMANT particle detector array. *Nucl. Instrum. Methods Phys. Res. A* **516,** 502–510 (2004).
18. Skeppstedt, Ö. *et al.* The EUROBALL neutron wall design and performance tests of neutron detectors. *Nucl. Instrum. Methods Phys. Res. A* **421,** 531–541 (1999).
19. Azaiez, F. EXOGAM: A γ-ray spectrometer for radioactive beams. *Nucl. Phys. A* **654,** 1003c–1008c (1999).
20. Simpson, J. *et al.* The EXOGAM array: a radioactive beam gamma-ray spectrometer. *Heavy Ion Phys.* **11,** 159–188 (2000); see also ⟨http://pro.ganil-spiral2.eu/laboratory/detectors/exogam⟩.
21. O'Leary, C. D. *et al.* Neutron-proton pairing, Coulomb effects and shape coexistence in odd-odd N=Z ^{46}V. *Phys. Lett. B* **459,** 73–80 (1999).
22. Lenzi, S. M. *et al.* Band termination in the N=Z odd-odd nucleus ^{46}V. *Phys. Rev. C* **60,** 021303(R) (1999).
23. Heese, J. *et al.* High spin states and shell model description of the neutron deficient nuclei ^{90}Ru and ^{91}Ru. *Phys. Rev. C* **49,** 1896–1903 (1994).
24. Mărginean, N. *et al.* Identification of excited states and shell model description of the N=Z+1 nucleus ^{91}Rh. *Phys. Rev. C* **72,** 014302 (2005).

Author Contributions G.d.F., S.B., E.C. and A.D. were responsible for setting up the EXOGAM germanium detectors and most of the related electronics and data acquisition system. J.N. was in charge of setting up the Neutron Wall detector system with its electronics. B.M.N., A.A., J.G., K.J., G.K., J.M., J.-N.S. and J.T. were responsible for the operation of the DIAMANT charged particle detector system with its associated electronics. All authors except J.B., J.G., K.J., G.K., R.L. and C.Q. participated in the measurements. B.C., F.G.M., T.B., K.A., K.L., E.C., G.d.F., A.Jo., M.P., R.W. and H.A.-A. performed the data analysis or were involved in different aspects of it. J.B. proposed the theoretical interpretation. The shell model calculations were performed by C.Q., R.L. and J.B. The manuscript text was prepared by B.C., with contributions from R.W., J.B., T.B., A.Jo., R.L., C.Q., A.A., G.d.A., E.C., Zs.D., F.G.M., A.Ju., S.M.L., R.M., B.M.N., J.N., M.P., F.R., D.S., M.S. and J.T. T.B. prepared the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.C. (cederwall@nuclear.kth.se).

LETTER

Vernier templating and synthesis of a 12-porphyrin nano-ring

Melanie C. O'Sullivan[1]*, Johannes K. Sprafke[1]*, Dmitry V. Kondratuk[1], Corentin Rinfray[1], Timothy D. W. Claridge[1], Alex Saywell[2], Matthew O. Blunt[2], James N. O'Shea[2], Peter H. Beton[2], Marc Malfois[3] & Harry L. Anderson[1]

Templates are widely used to arrange molecular components so they can be covalently linked into complex molecules that are not readily accessible by classical synthetic methods[1–7]. Nature uses sophisticated templates such as the ribosome, whereas chemists use simple ions or small molecules. But as we tackle the synthesis of larger targets, we require larger templates—which themselves become synthetically challenging. Here we show that Vernier complexes can solve this problem: if the number of binding sites on the template, n_T, is not a multiple of the number of binding sites on the molecular building blocks, n_B, then small templates can direct the assembly of relatively large Vernier complexes where the number of binding sites in the product, n_P, is the lowest common multiple of n_B and n_T (refs 8, 9). We illustrate the value of this concept for the covalent synthesis of challenging targets by using a simple six-site template to direct the synthesis of a 12-porphyrin nano-ring with a diameter of 4.7 nm, thus establishing Vernier templating as a powerful new strategy for the synthesis of large monodisperse macromolecules.

The principle behind all Vernier systems is the formation of a moiré pattern. Moiré patterns are observed on a wide range of length scales whenever two non-commensurate periodicities are superimposed. In the simplest case, two sets of parallel lines, with different spacings, produce a striped pattern with a periodicity that is the lowest common multiple of those of the two components, like the beat pattern generated by adding two sine waves of slightly different wavelengths (Fig. 1a). Engineers use moiré patterns and Vernier scales to amplify small mismatches in separation that would otherwise be difficult to resolve. Similarly in chemistry, the formation of a Vernier complex between components with different numbers of binding sites provides a way to amplify the molecular length scale, generating precisely defined assemblies (Fig. 1b). Although the idea of molecular Vernier systems has been discussed for more than 20 years[10], only two examples have been investigated experimentally[8,9]. The idea that the formation of Vernier complexes between a template and a molecular building block can serve as a powerful strategy for the synthesis of large macro-cycles (Fig. 1c) is based on the realization that Vernier complexes do not need to be linear. A template with six binding sites, T6, binds to a building block with four binding sites, l-P4, to form the Vernier complex (l-P4)$_3$•(T6)$_2$, which probably consists of two isomers. The l-P4 units can then be coupled together covalently to give a figure-of-eight complex, c-P12•(T6)$_2$. Displacement of the template gives the free macro-cycle c-P12. Vernier templating allows simple templates to be used to direct the formation of complex monodisperse targets.

We have demonstrated the concept of Vernier templating (Fig. 1c) by synthesizing a butadiyne-linked, π-conjugated 12-porphyrin nano-ring (Fig. 2a). Palladium-catalysed oxidative coupling of the linear porphyrin tetramer l-P4 in the presence of the hexapyridyl template T6 gave the figure-of-eight complex c-P12•(T6)$_2$ in 39% isolated yield (the only other products are insoluble polymers). Treatment of this

complex with an excess of pyridine, as a competing ligand, resulted in quantitative conversion to the free 12-porphyrin nano-ring c-P12 (isolated in 96% yield). Both c-P12•(T6)$_2$ and c-P12 were fully characterized by ^{1}H NMR and matrix-assisted laser desorption/ionization mass spectrometry (Supplementary Information). The ^{1}H NMR spectrum of c-P12•(T6)$_2$ (Fig. 2b, top), is fully consistent with the expected D_2 symmetry, with 24 β-pyrrole doublets and 12 t-butyl singlets. The resonances from the porphyrin units near the central crossing point are spread over a wide range of chemical shifts. We have assigned most of these resonances using a variety of two-dimensional NMR techniques, and observed nuclear Overhauser enhancements between resonances of the proximate porphyrin units at the centre of the structure (Supplementary Fig. 20). The ^{1}H NMR spectrum of c-P12 is extremely simple, with just two sharp β-pyrrole doublets, two aromatic signals and one t-butyl singlet, demonstrating its D_{12h} symmetry. The structures of c-P12•(T6)$_2$ and c-P12 are supported by solution-phase small-angle X-ray scattering (SAXS) data. Recently this technique has emerged as a valuable method for characterizing synthetic supramolecular architectures[11]. The experimental SAXS electron density pair-distribution functions, obtained from dilute solutions of c-P12•(T6)$_2$ and c-P12 in toluene, match the simulated pair-distribution functions for calculated geometries (Fig. 3a, b). The free nano-ring, c-P12, is quite flexible in
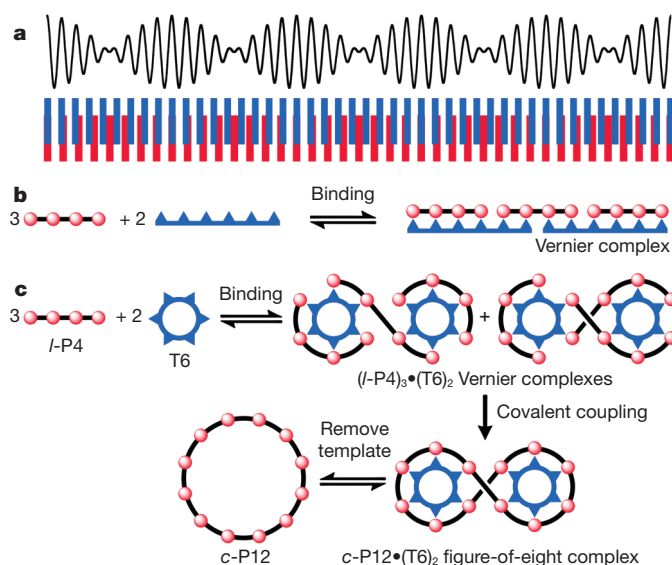

Figure 1 | Vernier templating. a, The basis of a Vernier measuring device is illustrated by the moiré pattern arising from the summation of two sine waves, or two stripes, with a periodicity ratio of 8:9. **b,** Formation of a molecular 3:2 Vernier complex. **c,** Vernier templating: the use of Vernier complex formation to direct the formation of a 12-site macrocycle, c-P12, using a 6-site template, T6.

[1]Department of Chemistry, Oxford University, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK. [2]School of Physics & Astronomy, University of Nottingham, Nottingham, NG7 2RD, UK. [3]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK.
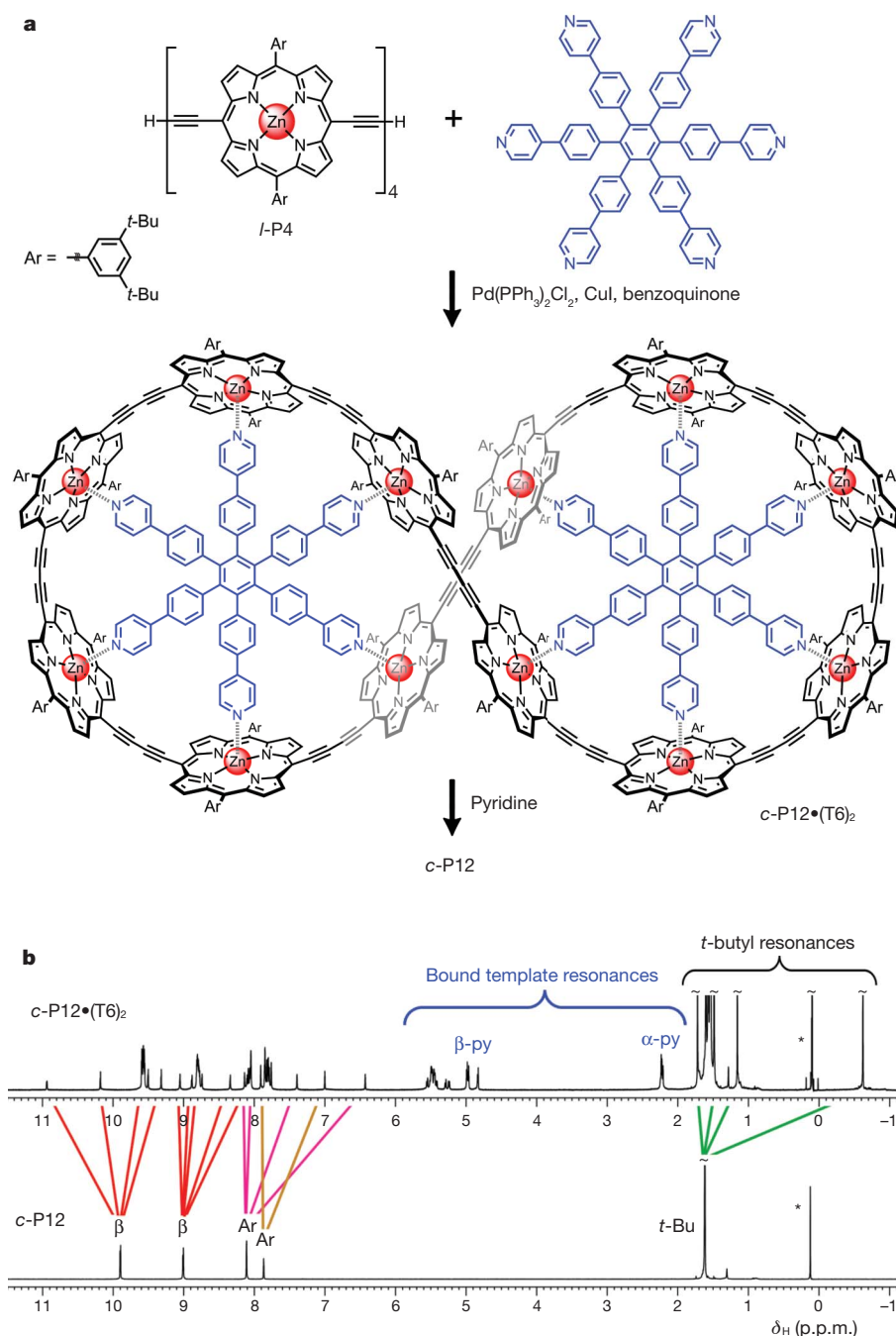*These authors contributed equally to this work.

Figure 2 | Synthesis and NMR characterization of the 12-porphyrin nano-ring. a, Vernier-templated synthesis of *c*-P12 through the formation of a figure-of-eight complex, *c*-P12•(T6)$_2$. **b**, ^{1}H NMR spectra of *c*-P12•(T6)$_2$ in CDCl$_3$ and *c*-P12 in CDCl$_3$/C$_5$D$_5$N. The coloured lines show how the spectrum of *c*-P12•(T6)$_2$ collapses on removal of the template (700 MHz, 298 K; diffusion-edited to exclude solvent signals; *silicon grease impurity; Ar, aryl; β, β-pyrrole; Bu, butyl; δ$_H$, chemical shift; py, pyridine).

solution and its SAXS data could only be adequately simulated by using a combination of several elliptical conformations. The structure of the 12-porphyrin nano-ring was also confirmed by scanning tunnelling microscopy (STM). Molecules were deposited using an electrospray source, on a Au(111) surface in ultrahigh vacuum[12]. In Fig. 3c, several nearly circular *c*-P12 molecules are shown adsorbed at a gold step edge (the line running diagonally across the image); in each case, it is possible to count the 12 porphyrin subunits (see also Supplementary Figs 39 and 40).

We also synthesized the nano-ring *c*-P12 by a 'classical' template-directed route (Fig. 4). Palladium-catalysed oxidative coupling of the linear porphyrin tetramer *l*-P4 in the presence of the dodecapyridyl template T12 gave *c*-P12•T12 in 35% isolated yield; treatment of this complex with pyridine results in quantitative conversion to the

free 12-porphyrin nano-ring *c*-P12. The efficiency of this classical templating route is similar to that of the Vernier route. However, it is more difficult to purify samples of *c*-P12 prepared using T12, and the T12 template is only available in small amounts from a low-yielding ten-step synthesis, whereas T6 is accessible in two steps from commercial materials[6]. This illustrates the tremendous advantage of Vernier templating as a way to gain access to large, complex targets using readily available templates.

Control experiments showed that oxidative coupling of *l*-P4 in the absence of a template gives linear polymers without forming any detectable *c*-P12. Coupling of the corresponding porphyrin monomer and dimer (*l*-P1 and *l*-P2) in the presence of template T6 yields the cyclic hexamer *c*-P6 together with small amounts of *c*-P12 (Supplementary
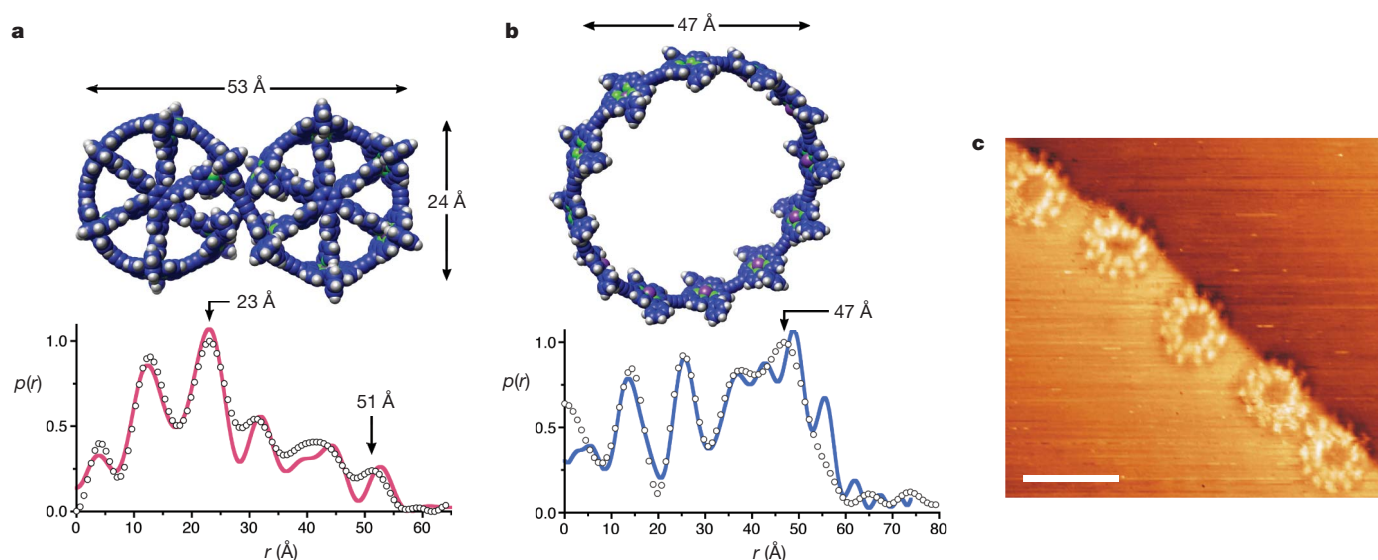
Figure 3 | Characterization of the 12-porphyrin nano-ring by SAXS and STM. a, b, SAXS pair-distribution data for *c*-P12•(T6)₂ in toluene (**a**) and *c*-P12 in toluene–pyridine (**b**); experimental points and simulated curves from the calculated structures are shown. *p*(*r*), pair-distribution function for electron density at separation *r*. **c**, STM image of *c*-P12 on a gold surface. The version of *c*-P12 used in this STM experiment had octyloxy side chains instead of *t*-butyls. STM scanning parameters: sample voltage, −1.8 V; tunnel current, 30 pA. Scale bar, 50 Å.
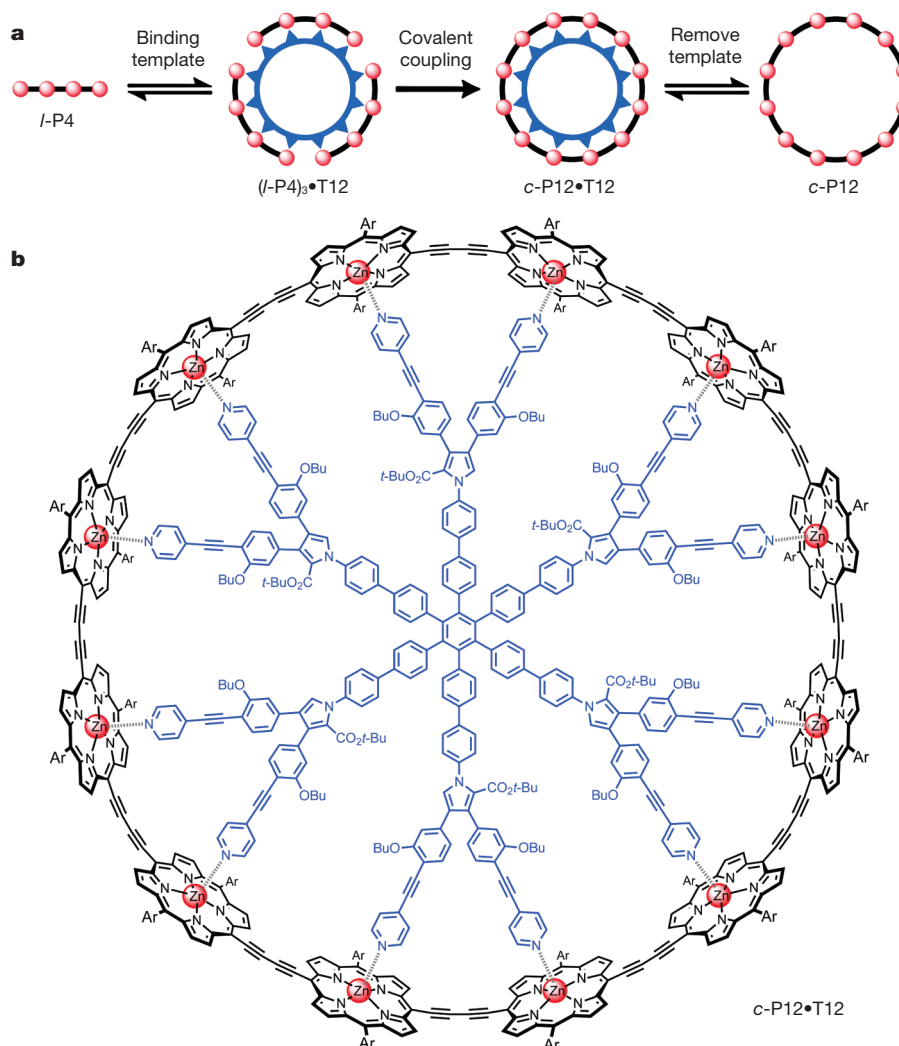
Figure 4 | Classical template-directed synthesis of nano-ring *c*-P12. a, Palladium-catalysed oxidative coupling of *l*-P4 in the presence of T12, followed by treatment with pyridine, yields *c*-P12. **b**, Structure of *c*-P12•T12 (Ar = 3,5-bis(*t*-butyl)phenyl).

Table 1); the yields of *c*-P12 from these routes are much lower than from Vernier templating.

The 12-porphyrin nano-ring *c*-P12, with a diameter of 4.7 nm, is among the largest π-conjugated macrocycles ever synthesized[5–7,13–15]. It will be interesting to investigate whether molecules of this type can support persistent ring currents analogous to those observed in mesoscopic metal rings[13,16]. Nano-rings of chromophores such as *c*-P12 are also a focus of attention because of their resemblance to natural light-harvesting chlorophyll complexes[11,17–21].

The results reported here indicate that Vernier templating will provide access to even larger cyclic chromophore arrays. For example, coupling a five-porphyrin building block, *l*-P5[22], in the presence of T6 should give *c*-P30, by way of a Vernier complex of stoichiometry $(l\text{-P5})_6\bullet(\text{T6})_5$. Formation of a stable Vernier complex is not essential for Vernier templating. The main role of the template may be to favour cyclization when a linear chain intermediate is produced with a number of binding sites, n_P, which is the lowest common multiple of n_B and n_T. Thus, in the Vernier synthesis of *c*-P12 (Figs 1c and 2a), the template could work simply by binding *l*-P12 (formed by oligomerization of *l*-P4) in such a way as to accelerate its cyclization. Vernier templating should be applicable to any template-directed cyclization reaction in which the components have well-defined binding sites, such that a precise mismatch can be set up between n_B and n_T. This strategy seems to be a general approach to the synthesis of monodisperse macromolecules of a size not previously accessible.

METHODS SUMMARY

Vernier-templated synthesis of *c*-P12•(T6)₂. The template T6 (19.5 mg, 19.6 μmol) and the linear porphyrin tetramer *l*-P4 (61.0 mg, 19.1 μmol) were dissolved in $CHCl_3$ (88 ml) by sonication for 1 h (bath sonicator). A catalyst solution was prepared by dissolving $Pd(PPh_3)_2Cl_2$ (17.7 mg, 25.2 μmol), Cu(ɪ) iodide (24.2 mg, 0.127 mmol) and 1,4-benzoquinone (56.1 mg, 0.52 mmol) in $CHCl_3$ (12 ml) and diisopropylamine (610 μl). The catalyst solution was added to the solution of *l*-P4 and T6, stirred for 1 h at 20 °C and then stirred for 1.5 h at 50 °C. The mixture was passed through a plug of alumina using $CHCl_3$ as eluent, and purified by size-exclusion chromatography (Biobeads SX-1, toluene). Recrystallization by layer addition of MeOH to a solution in CH_2Cl_2 yielded *c*-P12•(T6)₂ as a dark-brown solid (29.0 mg, 39%). See Supplementary Information for full details of the characterization of *c*-P12•(T6)₂ and related compounds.

Synchrotron radiation SAXS. A solution of the sample in toluene, or toluene and 1% pyridine (concentration, ~0.1 mM) was placed in a cell with mica windows. Data were collected on beamline I22 at the Diamond Light Source (UK), with the detector at a distance of 1.25 m from the sample cell; the momentum transfer range was $0.03 \text{ Å}^{-1} < q < 1.0 \text{ Å}^{-1}$ ($q = 4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 1.00 \text{ Å}$ is the X-ray wavelength). The data were normalized to the intensity of the incident beam, and scattering due to the solvent was subtracted.

Scanning tunnelling microscopy. A solution of *c*-P12 (with octyloxy side chains) in toluene–methanol (3:1 by volume, with 5% pyridine; concentration, 200 μg ml⁻¹) was deposited on the substrate (gold on mica) by electrospray. Images were acquired in ultrahigh vacuum using electrochemically etched tungsten tips, in constant-current mode at 20 °C.

1. Diederich, F. & Stang, P. J. (eds) *Templated Organic Synthesis* (VCH-Wiley, 1999).
2. Anderson, S., Anderson, H. L. & Sanders, J. K. M. Expanding roles for templates in synthesis. *Acc. Chem. Res.* **26,** 469–475 (1993).
3. Chichak, K. S. *et al.* Molecular Borromean rings. *Science* **304,** 1308–1312 (2004).
4. Crowley, J. D. *et al.* An unusual nickel-copper-mediated alkyne homocoupling reaction for the active-template synthesis of [2]rotaxanes. *J. Am. Chem. Soc.* **132,** 6243–6248 (2010).
5. Jung, S.-H. *et al.* A conjugated polycarbazole ring around a porphyrin. *Angew. Chem. Int. Ed.* **45,** 4685–4690 (2006).
6. Hoffmann, M. *et al.* Enhanced π-conjugation around a porphyrin[6] nanoring. *Angew. Chem. Int. Ed.* **47,** 4993–4996 (2008).
7. Hoffmann, M., Wilson, C. J., Odell, B. & Anderson, H. L. Template-directed synthesis of a π-conjugated porphyrin nanoring. *Angew. Chem. Int. Ed.* **46,** 3122–3125 (2007).
8. Kelly, T. R., Xie, R. L., Weinreb, C. K. & Bregant, T. A molecular vernier. *Tetrahedr. Lett.* **39,** 3675–3678 (1998).
9. Hunter, C. A. & Tomas, S. Accurate length control of supramolecular oligomerization: Vernier assemblies. *J. Am. Chem. Soc.* **128,** 8975–8979 (2006).
10. Lindsey, J. S. Self-assembly in synthetic routes to molecular devices. Biological principles and chemical perspectives: a review. *N. J. Chem.* **15,** 153–180 (1991).
11. Kelly, R. F. *et al.* Intramolecular energy transfer within butadiyne-linked chlorophyll and porphyrin dimer-faced, self-assembled prisms. *J. Am. Chem. Soc.* **130,** 4277–4284 (2008).
12. Saywell, A. *et al.* Conformation and packing of porphyrin polymer chains deposited using electrospray on a gold surface. *Angew. Chem. Int. Ed.* **49,** 9136–9139 (2010).
13. Mayor, M. & Didschies, C. A giant conjugated molecular ring. *Angew. Chem. Int. Ed.* **42,** 3176–3179 (2003).
14. Zhang, F., Götz, G. & Winkler, H. D. F. Schalley, C. A. & Bäuerle, P. Giant cyclo[*n*]thiophenes with extended π-conjugation. *Angew. Chem. Int. Ed.* **48,** 6632–6635 (2009).
15. Nakao, K. *et al.* Giant macrocycles composed of thiophene, acetylene, and ethylene building blocks. *J. Am. Chem. Soc.* **128,** 16740–16747 (2006).
16. Bleszynski-Jayich, A. C. *et al.* Persistent currents in normal metal rings. *Science* **326,** 272–275 (2009).
17. McDermott, G. *et al.* Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature* **374,** 517–521 (1995).
18. Roszak, A. W. *et al.* Crystal structure of the RC-LH1 core complex from *Rhodopseudomonas palustris. Science* **302,** 1969–1972 (2003).
19. Nakamura, Y., Aratani, N. & Osuka, A. Cyclic porphyrin arrays as artificial photosynthetic antenna: synthesis and excitation energy transfer. *Chem. Soc. Rev.* **36,** 831–845 (2007).
20. Hajjaj, F. *et al.* Assemblies of supramolecular porphyrin dimers in pentagonal and hexagonal arrays exhibiting light-harvesting antenna function. *J. Am. Chem. Soc.* **128,** 4612–4623 (2006).
21. Chang, M.-H., Hoffmann, M., Anderson, H. L. & Herz, L. M. Dynamics of excited-state conformational relaxation and electronic delocalization in conjugated porphyrin oligomers. *J. Am. Chem. Soc.* **130,** 10171–10178 (2008).
22. Taylor, P. N. & Anderson, H. L. Cooperative self-assembly of double-strand conjugated porphyrin ladders. *J. Am. Chem. Soc.* **121,** 11538–11545 (1999).

Author Contributions H.L.A. designed the project and wrote the manuscript. J.K.S., M.C.O., D.K. and C.R. carried out most of the experimental work. T.D.W.C. provided expertise with NMR analysis. STM was performed by A.S. and M.O.B., supervised by J.N.O. and P.H.B. SAXS analysis was performed by J.K.S. and H.L.A. with help from M.M. All authors edited the manuscript.

LETTER

Sensing the anomeric effect in a solvent–free environment

Emilio J. Cocinero[1]†, Pierre Çarçabal[1]†, Timothy D. Vaden[1]†, John P. Simons[1] & Benjamin G. Davis[2]

The anomeric effect is a chemical phenomenon[1–9] that refers to an observed stabilization[10] of six-membered carbohydrate rings when they contain an electronegative substituent at the C1 position of the ring. This stereoelectronic effect influences the three-dimensional shapes of many biological molecules. It can be manifested not only in this classical manner involving interaction of the endocyclic oxygen atom (O5) found in such sugars with the C1 substituent (endo-anomeric effect) but also through a corresponding inter-action of the electronegative exocyclic substituent with O5 (exo-anomeric effect). However, the underlying physical origin(s) of this phenomenon is still not clear[1,3,4,11–14]. Here we show, using a com-bination of laser spectroscopy and computational analysis, that a truncated peptide motif can engage the two anomers of an isolated sugar in the gas phase, an environment lacking extraneous factors which could confound the analysis. (Anomers are isomers that dif-fer in the orientation of the substituent at C1.) Complexes formed between the peptide and the α- or β-anomers of D-galactose are nearly identical structurally; however, the strength of the polariza-tion of their interactions with the peptide differs greatly. Natural bond order calculations support this observation, and together they reveal the dominance of the exo- over the endo-anomeric effect. As interactions between oxygen atoms at positions C1 and C2 (O1 and O2, respectively) on the pyranose ring can alter the exo/endo ratio of a carbohydrate, our results suggest that it will be important to re-evaluate the influence, and biological effects, of substituents at position C2 in sugars.

The preferential stabilization[10] of pyranose sugar rings when they contain an axial electronegative substituent at C1 (Fig. 1a) is contrary
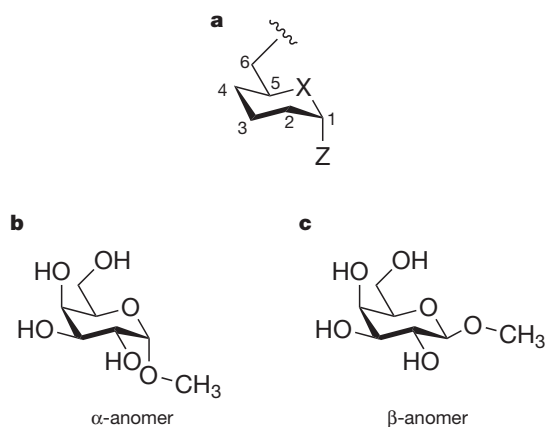
to expectations based on considerations of steric or solvation factors[15]. The balance between these factors, and stereoelectronic factors pro-vided by the anomeric effect, has made this effect one of the most celebrated structural theories in chemistry. Since its identification in 1955[1], the generalized implications of the anomeric effect have spread much further, leading to successful interpretations of other well-known structural phenomena, such as the *gauche* effect[15]. As a result, the anomeric effect has been used to interpret widespread stabilizing and orbital interaction effects well beyond those found in carbohydrates, in both cyclic and acyclic molecules, especially those that contain the general motif R-X-CH₂-Z where Z is an atom of an electronegative element and X is typically either an O or S atom.

Despite its importance in both chemistry and biology, a clear dissec-tion of the anomeric effect in archetypal, representative molecules—such as biologically relevant carbohydrates—has not been possible. This highlights the difficulty of generating such molecules in isolated (unsolvated) conditions and *a fortiori* the paucity of methods that are able to discern underlying effects and parameters in such cases. An ideal scenario would involve the isolation of carbohydrate mol-ecules in the gas phase, free from the influence of solvents, and their interrogation using some form of relevant spectroscopy/analytical method and/or sensor. We describe here a method that allows the creation and isolation of a peptidic sensor, capable of binding the archetypal carbohydrate, D-galactose (widely found in biology) in two different anomeric forms, α and β (Fig. 1b and c).

N-Acetyl-L-phenylalanine methylamide, chosen as a conforma-tionally flexible peptidic sensor (Fig. 2), has amide groups at both its *N*- and *C*-termini. These provide hydrogen bond donor and acceptor motifs with the potential (in the extended peptide conformation[16]) to engage in a pincer-like coordination to the carbohydrate. Gas phase complexes of the sensor with the carbohydrate anomers were created by laser ablation from a thin film of a powdered carbohydrate–peptide–graphite mixture, into an expanding argon jet which passed through a skimmer to create a collimated beam of cold molecular complexes. Their ionization through resonant two-photon ultraviolet excitation followed by time-of-flight mass spectrometry provided mass-resolved electronic excitation spectra (Supplementary Fig. 1). Conformation-specific vibrational spectra could then be recorded in the key O–H and N–H bond stretch regions (which are extraordinarily sensitive to hydrogen-bonded interactions) through infrared ion-dip double resonance spectroscopy[17]. Spectral analyses and structural assignments (Supplementary Figs 2, 3; Supplementary Table 1) were based on computational calculations, using density functional theory (DFT). An iterative approach was followed to achieve the best accord with the experimental spectra (Supplementary Information). In this way, calculations and experimentally determined data converged to provide precise structural information. The structures that corre-sponded best with experiment (Fig. 3) also presented the lowest cal-culated relative energies.

Figure 1 | Structures of sugars. a, A generalized structure of a sugar that displays anomeric effects, where typically X = O or S, and Z is an electron withdrawing (electronegative) substituent; in most natural sugar structures X = Z = O. **b**, **c**, Schematic structures of the α-anomer (**b**) and the β-anomer (**c**) of methyl D-galactopyranoside.

[1]Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ UK. [2]Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK. †Present addresses: Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV-EHU), Apartado 644, E-48940, Bilbao, Spain (E.J.C.); Institut des Sciences Moléculaires d'Orsay – CNRS, Université Paris Sud, F91405 Orsay Cedex, France (P.C.); Department of Chemistry and Biochemistry, Rowan University, 201 Mullica Hill Road, Glassboro, New Jersey 08028, USA (T.D.V.).
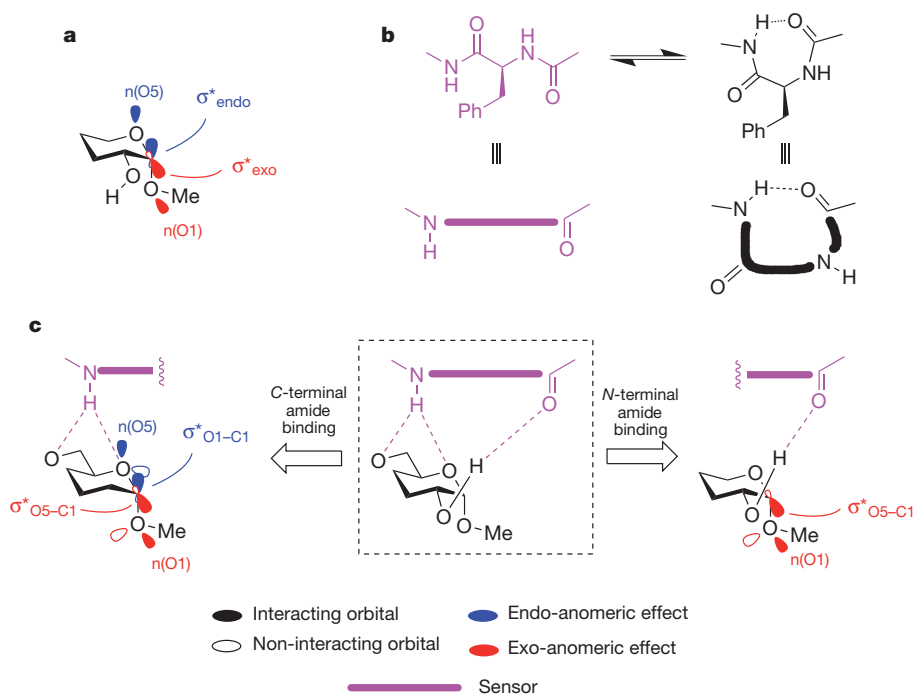
Figure 2 | The anomeric effect and the design of a sensor to isolate and detect it. a, Participating electron orbitals in the pyranose sugar structure. Exo, n(O1)→σ*(C1–O5) (red), and endo, n(O5)→σ*(C1–O1) (blue) hyperconjugative lone pair(n)–antibonding orbital(σ*) interactions contribute to the effect. **b**, A peptidic sensor to bind sugars. *N*-Acetyl-L-phenylalanine-methylamide has *N*- and *C*-terminal amides. It displays two primary conformers that are in equilibrium with each other: extended (shown in purple) and cyclical-γ-turn. These are represented in **c** using the 'cartoon' representations shown, as indicated by the equivalence signs. **c**, In extended form, the sensor simultaneously engages O5 and O6 with its *N*-terminal-amide and OH2 with its *C*-terminal-amide (centre). This is a combination of two binding modes. First (right), the *C*-terminal-amide engages OH2 and so prevents masking of the exo-anomeric effect, which is mediated by the orbitals shown in red. Second (left), the *N*-terminal-amide is sensitive to electron density on O5, providing a 'readout' of anomeric effects.

Two carbohydrate–peptide complexes were observed in our experiments; these gave different mass-resolved electronic spectra, making possible studies of their respective infrared spectra (Fig. 3). The two complexes, containing α- and β-sugars, although nearly identical in their structural composition, present significantly different intermolecular bonding properties, reflected in some striking differences in these spectra (Fig. 3). On moving from the β- to the α-complex, the (*C*-terminal) N–H vibrational band in the sensor is displaced by ~40 cm^{-1} to a lower wavenumber, which indicates a strengthening of its hydrogen bonding to the carbohydrate. At the same time, the OH2 band shifts by ~80 cm^{-1} to a higher wavenumber, indicating a relative weakening in the strength of its hydrogen bond to the (*N*-terminal) C=O group. In contrast, the displacement of bands associated with the non-participatory hydroxyl groups at OH6, OH4 and OH3 (and also the non-interacting (*N*-terminal) NH group on the peptide) remain in much the same position in both spectra. The computed structures that provide the best fit to the experimental spectra indicate a contraction in the NH•••O6 distance on moving from the β- to the α-anomeric complex, accompanied by an extension of the NH•••O5 and OH2•••O=C distances (see Fig. 4 and Table 1). The sensor engages selectively with O5 and OH2, sites relevant to the anomeric effect, and there is a subtle movement of the sensor towards O6 and away from OH2 and O5 (Fig. 4).

Different orientations of the OCH$_3$ group at the anomeric centre C1 cannot lead to additional hydrogen-bonds, and changes in the orientation of OH2 and the exocyclic hydroxymethyl group do not lead to significant differences in energy between the two anomers (Supplementary Figs 2, 3; Supplementary Table 2). Together, these suggest an interaction mediated by changes at O5, modulated indirectly by anomeric configuration. The truncated peptide motif has uniquely and discretely revealed a dramatic alteration of the bonding properties in the two anomers of methyl D-galactopyranoside, and infrared spectroscopy has provided an experimental signature of this single anomeric effect.

The anomeric modulation of O5 activity revealed through the spectral shifts and structural analysis prompted us to probe more deeply into its origin. A natural bond orbital analysis[11] (Table 1; Supplementary Tables 2–4) consistent with the observed experimental data and the structures described in Figs 3 and 4 also revealed key aspects of orbital changes. Endo- and exo-anomeric effects associated, respectively, with n(O5)→σ*(C1–O1) and n(O1)→σ*(C1–O5) hyperconjugative interactions, are present in each case but the exo-anomeric effect is the more dominant. Thus, electronic orbital occupancy for the σ*(C1–O1) orbital in the β-complex is substantially lowered to 0.034*e*, indicating a much stronger exo- than endo-anomeric effect. A smaller difference is also observed in the α-complex, where the corresponding orbital occupancies are 0.062*e* and 0.051*e*. The stronger exo-anomeric effect leads to a lengthening of the C1–O5 bond compared to the C1–O1 bond, and an increase in the second order perturbation energy, $\Delta E_{n\sigma^*}$. Although the exo-anomeric effect dominates overall, the magnitude of the endo-anomeric component is much greater in the α-anomer and the calculated lone pair electron occupancy on O5 decreases by ~0.05*e*. The resulting bond length data (Table 1) are consistent with bond length observations in crystals[5,7,18,19].

These changes correlate with those detected and confirmed by the spectroscopic behaviour of the sensor. The relative reduction in the 'lone pair' electron density at O5 in the α-anomer, caused by hyperconjugative interaction, leads to a weakening of the interaction between the sensor and O5, a consequence of the increase in the endo-anomeric component. It also changes the relative distances of the NH binding site from the O5 and O6 atoms, and there is a concomitant tightening of bonding to O6 and a consequent lengthening of the hydrogen bond OH2•••O=C. These global changes caused by the anomeric effect are reflected in shifts in the two key bond stretch frequencies in the gas phase infrared spectra (Fig. 3).

In summary, the peptide sensor complexes of methyl D-galactopyranoside display spectral properties that are a direct 'readout' of anomeric configuration. The DFT calculations and the natural
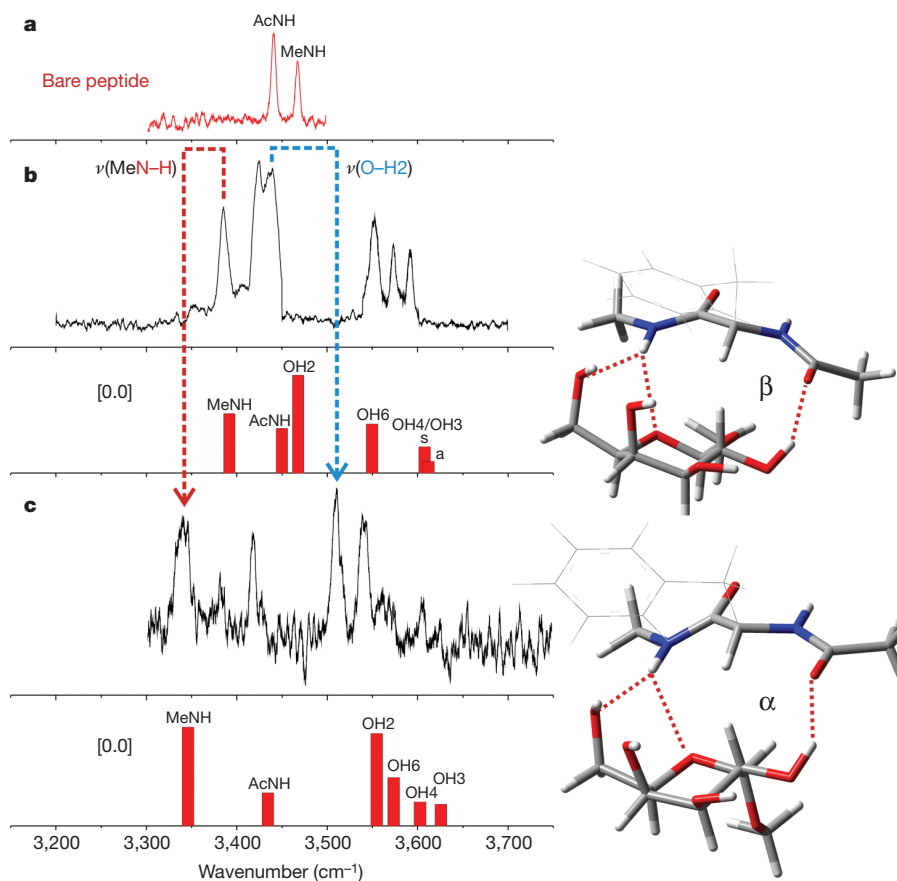
Figure 3 | Infrared spectra of the sensor and sugars in the gas phase, and determined structures. a–c, Experimental infrared ion-dip spectrum associated with the extended peptide sensor backbone alone (**a**), the β-complex (**b**) and the α-complex (**c**). Computed vibrational spectra associated with their lowest energy structures are shown underneath the experimental spectrum in **b** and **c**, with the corresponding determined structures shown in stick

representation to the right (phenylalanine benzyl substituent shown faded for clarity). Dashed red and blue arrows indicate 'red' and 'blue' shifts in the spectra; dotted red lines indicate key hydrogen-bonds; 'a' and 's' mark asymmetric and symmetric stretches, respectively; [0,0] indicates a structure corresponding to a global minimum energy.

bond orbital analysis based upon them, successfully rationalize the different vibrational signatures associated with the two anomeric complexes, isolated in the gas phase. The enhanced engagement of the O5 atom in the β-anomer strongly supports the suggestion of an increased electron density or polarization on O5 relative to the α-anomer, consistent with hyperconjugative effects. Furthermore, by engaging OH2 the truncated peptide sensor uniquely removes the strong potential influence of OH2–O1 interactions on the *anti/gauche* conformational equilibrium of O5-C1-O1-CH$_3$ (Fig. 2c). Their subsidiary influence

can dominate anomeric effects in some systems[20], masking the exo-anomeric effect and leading to an under-estimate of its relative importance. The present strategy has allowed the peptide sensor to probe the combined influence of intrinsic effects associated with the carbohydrate core structure and, to our knowledge for the first time, to reveal through experiment the physical mechanisms underlying the anomeric effect.

Of course, this analysis does not exclude the utility of other analytical interpretations in alternative environments[12,14,21,22]. When carbohydrates are placed in biological and chemical environments, anomeric effects can be attributed to a combination of potential influences: these include steric,
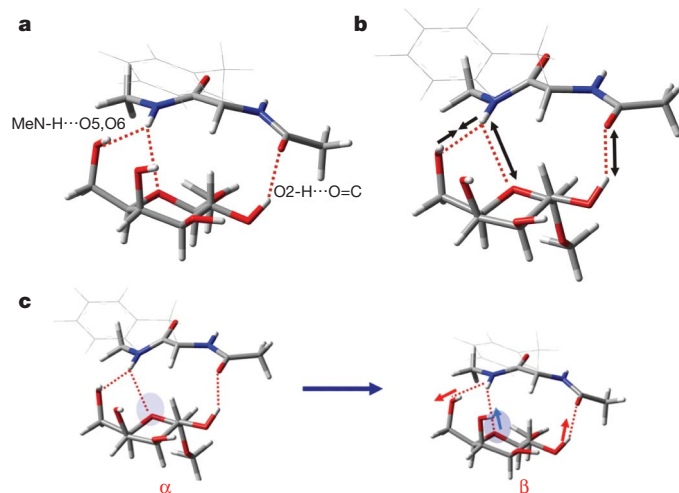
Figure 4 | Mode of action of the peptide sensor in detecting structural changes associated with the anomeric effect. a, In 'unfurled' linear conformation, the sensor uses its *C*-terminal-amide N–H hydrogen bond donor to engage O5,O6 and its *N*-terminal-amide carbonyl hydrogen bond acceptor to engage O2. The key hydrogen-bonds are shown as red dotted lines. **b,** The N–H and O–H stretch regions, which may be interrogated by infrared ion-dip double resonance spectroscopy, are highly sensitive to alterations of hydrogen-bonding; these can be used to directly probe the carbohydrate. These alterations are shown as black, opposed single-headed arrow pairs (shortened) and double-headed arrows (lengthened). Global changes caused by the anomeric effect are reflected in the shift in key bond stretch frequencies, including measurable electron density changes on O5 that can be directly 'sensed' by the peptide. **c,** The overall 'anomeric effect' detected is shown as a summary of observations. On moving from α- to β-anomer, the peptidic sensor changes reflect interacting electron density at O5 (highlighted grey region and blue arrow) as it moves away from O5 and OH2 and towards O6 (as shown by arrow directions).

Table 1 | Natural bond orbital analysis of the anomeric effect observed in the α and β carbohydrate–peptide complexes

Orbital hyperconjugation	$\Delta E_{n\sigma}*$ (kJ mol^{-1})		R(X–Y) (Å)†		$\sigma*_{occ}$ (e)		Comments
	α	β	α	β	α	β	
Endo							
n(O5)→σ*(C1–O1)	60.74	20.36	NA	NA	0.051	0.034	$\sigma*_{occ}$ (α>β)
R(C1–O1)†	NA	NA	1.395	1.382	NA	NA	$\delta R = -0.013$ Å
Exo							
n(O1)→σ*(C1–O5)	72.52	72.56	NA	NA	0.062	0.059	$\sigma*_{occ}$ (α≈β)
R(C1–O5)†	NA	NA	1.419	1.422	NA	NA	$\delta R \approx 0.003$ Å
Hydrogen bonding							
n(O6)→σ*(H–NCH$_3$)	46.36	26.00	NA	NA	0.038	0.035	NH→O6 (α>β)
R(O6•••H)†	NA	NA	1.99	2.14	NA	NA	$\delta R \approx 0.15$ Å
n(O5)→σ*(H–NCH$_3$)	4.51	9.41	NA	NA	0.038	0.035	NH→O5 v. weak (but β>α)
R(O5•••H)†	NA	NA	2.61	2.39	NA	NA	$\delta R \approx -0.22$ Å
n(O(=C))→σ*(H–O2)	29.55	30.68	NA	NA	0.021	0.027	OH2→O=C (β>α)
R((C=)O•••H)†	NA	NA	1.95	1.90	NA	NA	$\delta R \approx -0.05$ Å

The calculations were performed for the bound anomers of methyl D-galactopyranoside in conformational structures they adopt within the peptide complexes. $\Delta E_{n\sigma}*$ is the stabilization energy associated with hyperconjugation, calculated through second order perturbation theory; R(X–Y) is interatomic distance; $\sigma*_{occ}$ is electron occupancy of antibonding orbital. NA, not applicable.
† From DFT (MO5-2X/6-31+G**) calculations fitting the experimental IR spectra.

electrostatic and solvation effects, as well as hydrogen bonding and hyperconjugation. In the absence of all other factors, however, the exo-anomeric n(O1)→σ*(C1–O5) interaction is the most dominant, even in the α-anomer. Effects that fine tune or remove the exo-anomeric effect are probably more variable and therefore more important for functional analysis. They include repulsive effects of O2 upon O1 in *cis*-configured α-glycosides, and the role of OH2→O1 hydrogen bonding in changing the orientation of O1 lone pairs. Given the widespread role of glycosidic linkages in biology, these results suggest a physical basis for possible evolutionary benefits associated with the prevalent variation of substituents at C2 of sugars, rather than at other positions, which gives rise to biologically important sugars such as D-mannose, 2-acetamido-2-deoxy-D-glucose, 2-acetamido-2-deoxy-D-galactose and a wide variety of 2-deoxy sugars.

METHODS SUMMARY

Following their preparation, combined peptide and carbohydrate samples were transferred into a pulsed Ar jet using a pulsed laser desorption source, to form a molecular beam of peptide–carbohydrate dimers. Absorption of a resonant infrared laser pulse depleted their population and the transient depletion was detected mass spectrometrically, following ionization of the selected molecular species by resonant two-photon ultraviolet laser excitation. Vibrational action ('ion-dip') spectra of the cold populations of each of the selected molecular complexes, isolated and stabilized in the jet, were generated by scanning the infrared laser frequency. Assignment of the complex molecular structures was based on a theoretical analysis of the experimental spectra. It began with the generation of a large set of possible cluster structures generated by a molecular mechanics force field, using an iterative Monte Carlo search of the conformational landscape, followed by their statistical analysis and sorting into distinct conformational families. This was coupled with an iterative strategy in which the experimental data were used to guide subsequent calculations. After the initial survey, the lowest energy structures were re-optimized using density functional theory, which was also used to determine their harmonic vibrational frequencies, infrared intensities and zero point energy corrections.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Edward, J. T. Stability of glycosides to acid hydrolysis. *Chem.* Ind. 1102–1104 (1955).
2. Eliel, E. L. & Giza, C. A. Conformational analysis.17. 2-Alkoxy- and 2-alkylthiotetrahydropyrans and 2-alkoxy-1,3-dioxanes. Anomeric effect. *J. Org. Chem.* **33,** 3754–3758 (1968).
3. Romers, C., Altona, C., Buys, H. R. & Havinga, E. in *Topics in Stereochemistry* Vol. 4 (eds Eliel, E. L. & Allinger, N. L.) 39–97 (Wiley Interscience, 1969).
4. Lemieux, R. U. Effects of unshared pairs of electrons and their solvation on conformational equilibria. *Pure Appl. Chem.* **15,** 527–548 (1971).
5. Kirby, A. J. *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen* (Springer, 1983).
6. Deslongchamps, P. *Stereoelectronic Effects in Organic Chemistry* (Pergamon, 1983).
7. Box, V. G. S. The role of lone pair interactions in the chemistry of the monosaccharides. The anomeric effect. *Heterocycles* **31,** 1157–1181 (1990).
8. Juaristi, E. & Cuevas, G. Recent studies of the anomeric effect. *Tetrahedron* **48,** 5019–5087 (1992).
9. Thatcher, G. R. J. *The Anomeric Effect and Associated Stereoelectronic Effects* (ACS, 1993).
10. Jungius, C. L. Isomeric changes of some dextrose derivatives and the mutarotation of the sugars. *Z. Phys. Chem.* **52,** 97–108 (1905).
11. Salzner, U. & Schleyer, P. v. R. *Ab initio* examination of anomeric effects in tetrahydropyrans, 1,3-dioxanes, and glucose. *J. Org. Chem.* **59,** 2138–2155 (1994).
12. Vila, A. & Mosquera, R. A. Atoms in molecules interpretation of the anomeric effect in the O-C-O unit. *J. Comput. Chem.* **28,** 1516–1530 (2007).
13. Wiberg, K. B. & Murcko, M. A. Rotational barriers. 4. Dimethoxymethane — the anomeric effect revisited. *J. Am. Chem. Soc.* **111,** 4821–4828 (1989).
14. Mo, Y. Computational evidence that hyperconjugative interactions are not responsible for the anomeric effect. *Nature Chem.* **2,** 666–671 (2010).
15. Eliel, E. L. & Wilen, S. H. *Stereochemistry of Organic Compounds* (Wiley Interscience, 1994).
16. Gerhards, M., Unterberg, C., Gerlach, A. & Jansen, A. Beta-sheet model systems: structures and vibrations of Ac–Phe–NHMe and its dimer. *Phys. Chem. Chem. Phys.* **6,** 2682–2690 (2004).
17. Robertson, E. G. & Simons, J. P. Getting into shape: conformational and supramolecular landscapes in small biomolecules and their hydrated clusters. *Phys. Chem. Chem. Phys.* **3,** 1–18 (2001).
18. Briggs, A. J., Glenn, R., Jones, P. G., Kirby, A. J. & Ramaswamy, P. Bond length and reactivity. Stereoelectronic effects on bonding in acetals and glucosides. *J. Am. Chem. Soc.* **106,** 6200–6206 (1984).
19. Lemieux, R. U., Koto, S. & Voisin, D. in *Anomeric Effect, Origin and Consequences* (eds Szarek, W. A. & Horton, D.) 17–29 (ACS Symposium Series, American Chemical Society, 1979).
20. Cramer, C. J., Truhlar, D. G. & French, A. D. Exo-anomeric effects on energies and geometries of different conformations of glucose and related systems in the gas phase and aqueous solution. *Carbohydr. Res.* **298,** 1–14 (1997).
21. Takahashi, O. *et al.* The anomeric effect revisited. A possible role of the CH/n hydrogen bond. *Carbohydr. Res.* **342,** 1202–1209 (2007).
22. Gillespie, R. J., Robinson, E. A. & Pilmé, J. Ligand close packing, molecular compactness, the methyl tilt, molecular conformations, and a new model for the anomeric effect. *Chem. Eur. J.* **16,** 3663–3675 (2010).

Author Contributions E.J.C., P.C. and T.V.D. performed the experiments and computation. J.P.S. and B.G.D. designed the study, analysed the data and wrote the paper. All authors discussed the results and made comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.G.D. (Ben.Davis@chem.ox.ac.uk) and J.P.S. (John.Simons@chem.ox.ac.uk).

METHODS

Experimental. *N*-acetyl L-phenylalanine methylamide and the α- and β-anomers of methyl D-galactopyranoside were synthesized or obtained as commercial samples and were used without further purification. The peptide–carbohydrate complexes were generated in the gas phase using a combination of pulsed laser desorption and molecular beam procedures. Ground powder samples of the carbohydrate and the peptide (in a 5:1 molar ratio) were thoroughly mixed with graphite powder (20% graphite/80% organic, w/w), deposited as a thin uniform surface layer on a graphite substrate, and placed in a vacuum chamber close to and just below the exit of a pulsed, cylindrical nozzle expansion valve (0.8 mm diameter). The species desorbed from the surface were entrained and cooled in an expanding argon jet (~4 bar backing pressure) before passing into the detection chamber through a 2 mm diameter skimmer to create a collimated molecular beam. This was crossed by one or two tunable laser beams in the extraction region of a linear, time-of-flight mass spectrometer. Mass-resolved resonant two-colour photoionization spectra of selected molecular complexes were recorded using a frequency-doubled pulsed Nd:YAG-pumped dye laser (Sirah), operated at 10 Hz. Their conformer-specific vibrational spectra were subsequently recorded in the O–H and N–H stretch regions, through infrared ion dip double resonance spectroscopy[17] using ultraviolet radiation tuned onto selected resonant two-photon ionization (R2PI) absorption bands and tunable infrared radiation in the range 3,200–3,800 cm^{-1}. The ultraviolet laser was pulsed at 10 Hz but the infrared laser, fired after a delay of ~150 ns, was pulsed at 5 Hz, to allow subtraction of the background signals. The infrared radiation (line-width 2–3 cm^{-1}, 5 mJ per pulse) was provided by the idler output of an OPO/OPA laser system (LaserVision), pumped by a pulsed Nd/YAG laser (Continuum, Surelite II). Several spectra, typically ≥5, were recorded and averaged to achieve acceptable signal:noise levels.

Computational strategy. The structural search followed an iterative approach that began with a series of unrestricted surveys of the many possible complex structures until no additional new structures were obtained. These surveys, which generated 3,000–4,000 structures, were conducted using the Monte Carlo multiple minimization procedure and the large scale low frequency mode torsional sampling procedures implemented in the MacroModel software (MacroModel v.8.5.207, Schrödinger, LLC21). These initial sets of structures were grouped into families distinguished by their association with extended or folded peptide backbones and by the number of strong hydrogen bonded interactions they presented. Subsequent geometry optimization of the lower energy members of each family (typically ~50 structures with relative energies ≤15 kJ mol^{-1}) using density functional theory, the 'dispersion-sensitive' MO5-2X functional and a 6-31+G** basis set, implemented in the Gaussian03 suite of programs, led to a new set of relative energies (corrected for zero point energies), molecular structures and vibrational spectra which could then be compared with experiment. The predicted harmonic wavenumbers of the N–H and O–H stretch modes were scaled by the factor 0.9419 to accommodate anharmonicity and aid comparisons with the experiment. Depending on the level of agreement between the computed and experimental spectra, the structures which provided the closest levels of agreement, judged first by the degree of correspondence between the experimental and computed wavenumbers of each vibrational band, and second, by their calculated relative energies, were further optimized, guided by the experimental spectra. The process was repeated iteratively, to minimize their calculated relative energies and achieve the best accord with the experimental spectra.

LETTER

Geochemical evidence for widespread euxinia in the Later Cambrian ocean

Benjamin C. Gill[1]†, Timothy W. Lyons[1], Seth A. Young[2], Lee R. Kump[3], Andrew H. Knoll[4] & Matthew R. Saltzman[5]

Widespread anoxia in the ocean is frequently invoked as a primary driver of mass extinction as well as a long-term inhibitor of evolutionary radiation on early Earth. In recent biogeochemical studies it has been hypothesized that oxygen deficiency was widespread in subsurface water masses of later Cambrian oceans[1,2], possibly influencing evolutionary events during this time[1–3]. Physical evidence of widespread anoxia in Cambrian oceans has remained elusive and thus its potential relationship to the palaeontological record remains largely unexplored. Here we present sulphur isotope records from six globally distributed stratigraphic sections of later Cambrian marine rocks (about 499 million years old). We find a positive sulphur isotope excursion in phase with the Steptoean Positive Carbon Isotope Excursion (SPICE), a large and rapid excursion in the marine carbon isotope record, which is thought to be indicative of a global carbon cycle perturbation[4,5]. Numerical box modelling of the paired carbon sulphur isotope data indicates that these isotope shifts reflect transient increases in the burial of organic carbon and pyrite sulphur in sediments deposited under large-scale anoxic and sulphidic (euxinic) conditions. Independently, molybdenum abundances in a coeval black shale point convincingly to the transient spread of anoxia. These results identify the SPICE interval as the best characterized ocean anoxic event in the pre-Mesozoic ocean and an extreme example of oxygen deficiency in the later Cambrian ocean. Thus, a redox structure similar to those in Proterozoic oceans[6–8] may have persisted or returned in the oceans of the early Phanerozoic eon. Indeed, the environmental challenges presented by widespread anoxia may have been a prevalent if not dominant influence on animal evolution in Cambrian oceans.

Cambrian carbonates preserve large, rapid (of a few million years duration or less) and globally correlated excursions in the marine carbon isotope record ($\delta^{13}C_{carb}$), indicating perturbations in the global carbon cycle[4,9–11]. The driving mechanisms behind these events, however, are poorly understood. What makes these excursions particularly interesting to geobiologists is that many of them coincide with biological events recorded by fossils, suggesting links between biological and environmental history[3,4]. Here we focus on the last large excursion of the period, the SPICE.

The SPICE is recorded as a +4–6‰ shift in $\delta^{13}C_{carb}$ that occurs globally in later Cambrian successions (beginning of the Furongian International Series and Paibian International Stage, 499 million years ago); it is thought to have lasted 2–4 million years[4,5]. A well-documented extinction of trilobites coincides with the onset of the SPICE on the palaeocontinent of Laurentia[4,12], and the isotopic excursion has also been correlated to intervals of biological turnover on other palaeocontinents[13]. The SPICE is also coincident with global changes in sea level; its onset coincides with a transgressive event, and its peak is concurrent with a lowstand recorded as the Sauk-II/III hiatus[4,5].

We report sulphur isotope data from six globally distributed stratigraphic sections across the SPICE event that reveal parallel, positive carbon and sulphur isotope excursions (Figs 1, 2 and 3). These sections represent diverse sedimentary environments. Thus, similarities among the trends, despite differences in depositional conditions, indicate the global and primary nature of the geochemical signals (see Supplementary Information for details of individual stratigraphic sections and data supporting the preservation of the geochemical signals).

The SPICE sulphur isotope excursion is one of the largest identified in the geologic record and is the first to be correlated globally at a high scale of resolution. This excursion occurs in both carbonate-associated sulphate ($\delta^{34}S_{CAS}$) and pyrite ($\delta^{34}S_{pyrite}$), which further supports a primary marine signal, and its magnitude indicates a major perturbation in the global sulphur cycle. There are, however, significant differences in the details of the sulphate sulphur isotope trends among basins. In particular, the pre-event $\delta^{34}S_{CAS}$ baseline differs in the various locations (Fig. 2): some records show relatively steady sulphur isotope values before the excursion (that is, western and eastern Laurentia), but Gondwanan data show a positive trend up-section before the excursion proper (Fig. 2).

Furthermore, despite overarching similarities, the absolute values and amplitudes of the excursion also differ among the studied basins. The Gondwanan record is the most extreme, with $\delta^{34}S_{CAS}$ values reaching almost +70‰ and an amplitude of +35‰ (Fig. 2). On the other end of the spectrum, the record in eastern Laurentia shows a peak value of +38‰ and amplitude of only +12‰ (Fig. 2). Sulphate isotope data from the contemporaneous Port au Port Group from Newfoundland show little evidence of a sulphur isotope excursion during the SPICE event[2], but those results contain only a single data point near the peak of the SPICE. The $\delta^{34}S_{CAS}$ relationships at this locality need to be investigated further before conclusions can be drawn.
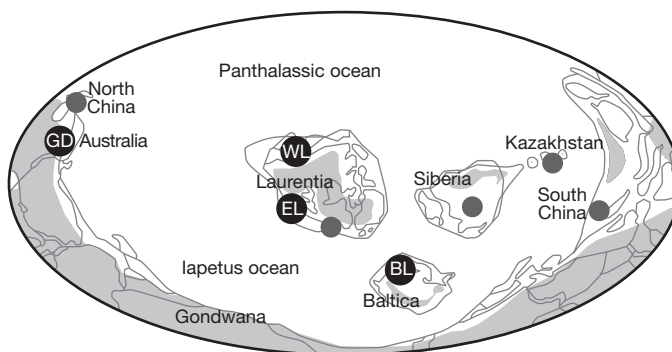

Figure 1 | Palaeo-reconstruction of the later Cambrian Earth. The reconstruction is based on ref. 27 and shows locations where the SPICE has been identified (filled circles). Locations investigated in this study are: Western Laurentia (WL) = Shingle Pass and Lawson Cove, Great Basin, USA; Eastern Laurentia (EL) = TE-1 Texas County Core, Missouri, USA; Gondwana (GD) = Mount Whelan no. 1 and Mount Murray, Queensland, Australia; Baltica (BL) = Andrarum no. 3 core, Sweden.

[1]Department of Earth Sciences, University of California, 900 University Avenue, Riverside, California 92521, USA. [2]Department of Geological Sciences, Indiana University-Bloomington, 1001 East 10th Street, Bloomington, Indiana 47405-1405, USA. [3]Department of Geosciences, Penn State University, 503 Deike Building, University Park, Pennsylvania 16802, USA. [4]Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA. [5]School of Earth Science, The Ohio State University, 275 Mendenhall Laboratory, 125 South Oval Mall, Columbus, Ohio 43210, USA. †Present address: Department of Earth and Planetary Sciences Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA.
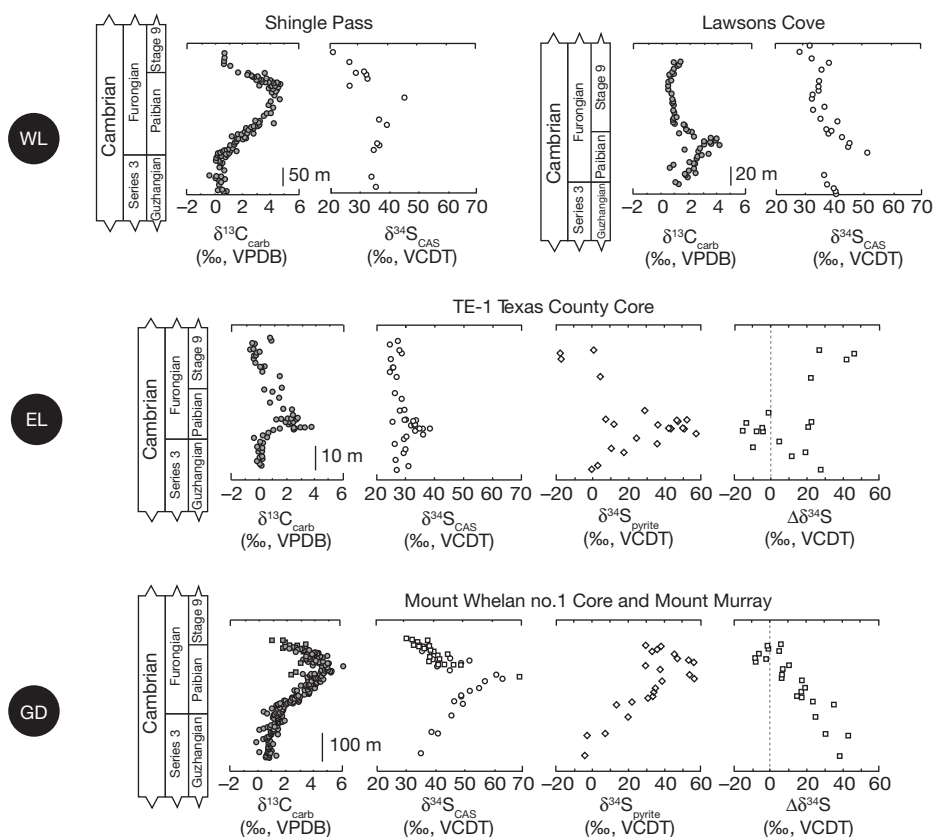
Figure 2 | Chemostratigraphies of the SPICE carbonate stratigraphic sections. Isotope data are plotted by stratigraphic height in metres. International series and stages are based on published biostratigraphy and the most recent definitions of the subdivisions for the Cambrian Period (see Supplementary Information). Carbon isotope profiles from Great Basin sections and Mount Whelan no. 1 core are from refs 10 and 4, respectively. CAS sulphur isotope profile for Shingle Pass is from ref. 15. On the GD $\delta^{13}C_{carb}$ and $\delta^{34}S_{CAS}$ plots, circles are from Mount Whelan no. 1 Core and squares from the Mount Murray outcrop. Carbon and sulphur isotope data are reported as per mil (‰) deviations from the isotope composition of Vienna Pee Dee Belemnite (VPDB) and Vienna Cañon Diablo Troilite (VCDT), respectively.

Overall, these isotopic differences support the idea that the sulphur reservoir in the later Cambrian ocean was spatially heterogeneous and that sulphate concentrations were therefore low[14,15]. We also observe that the sulphate isotope excursion peaks stratigraphically slightly before the carbon isotope maximum (see Supplementary Fig. 2), which suggests that the sulphate reservoir was relatively more sensitive to change than the marine pool of dissolved inorganic carbon. This state of sulphate in later Cambrian seawater differed greatly from the modern reservoir, which is relatively homogeneous globally with a concentration of 28 mmol kg^{-1} (mM) and a sulphur isotope composition of +21‰. This contrast with the modern value indicates that the residence time of sulphate in the Cambrian ocean was much less than today's.

The parallel behaviour between the carbon and sulphur isotope excursions (Fig. 2) suggests that the SPICE records a transient increase

in the amount of carbon and sulphur buried as organic matter and pyrite (FeS$_2$) in marine sediments. Such parallel burial occurs in anoxic marine sediments and beneath euxinic water columns[16]—that is, beneath water columns that are both anoxic and contain free hydrogen sulphide. Organic matter fuels microbial sulphate reduction, and pyrite is formed when H$_2$S produced from microbial sulphate reduction reacts with iron minerals and is buried along with the residual organic matter. Ultimately, the burial of both species results in the removal of carbon and sulphur from the ocean. This coupling can result in positive isotope shifts for both species in seawater: the carbon and sulphur leaving the ocean through burial are enriched in ^{12}C and ^{32}S via isotope fractionations accompanying photosynthetic and microbial sulphate reduction pathways, respectively, leaving the seawater correspondingly enriched in ^{13}C and ^{34}S.
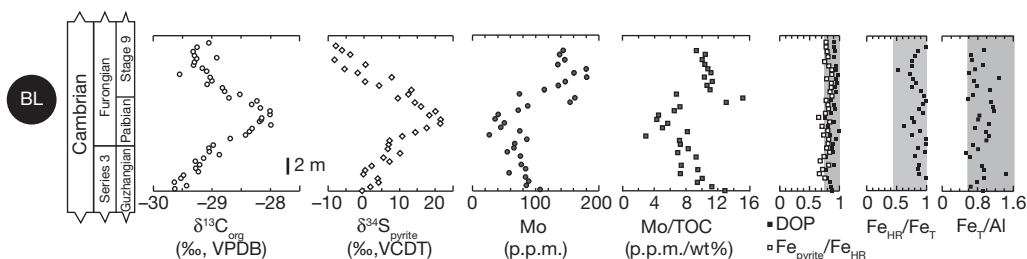

Figure 3 | Chemostratigraphy from the Andrarum no. 3 Core of the Alum Shale, Sweden (Baltica). The carbon isotope profile is from ref. 28. The shaded regions of the plots of degree of pyritization (DOP), Fe$_{pyrite}$/Fe$_{HR}$, Fe$_{HR}$/Fe$_T$ and Fe$_T$/Al cover values that indicate anoxia and euxinia: Fe$_T$/Al values appreciably above 0.5 and Fe$_{HR}$/Fe$_T$ above 0.4 indicate deposition under anoxic waters[29],

and DOP and Fe$_{pyrite}$/Fe$_{HR}$ values above 0.75 are conservatively diagnostic of euxinic environments[30]. We note that the decrease and minimum in Mo and Mo/TOC correspond to the initiation and peak of the carbon and sulphur isotope excursions, respectively. C$_{org}$, organic carbon. TOC, total organic carbon. Fe$_{pyrite}$, pyrite iron. Fe$_{HR}$, highly reactive iron. Fe$_T$, total iron.

We tested this hypothesis by modelling the ocean inventories of carbon and sulphur during the SPICE. Specifically, we constructed a simple box model that simulates the cycling of each element in the ocean (see Supplementary Information for details). The model shows that the isotope excursions can be replicated by transiently increasing the amount of organic carbon buried by factors of 1.5 to 2.5 and pyrite sulphur by factors of 2.5 to 4.5 for a duration of 0.5 to 1.5 million years (Fig. 4, and see Supplementary Information for additional model details).

Importantly, our model puts quantitative constraints on the size of the marine sulphate reservoir during the later Cambrian period. An assumption of pre-SPICE sulphate concentrations greater than 2.5 mM demands more than 8 million years for recovery of $\delta^{34}S_{sulphate}$ (that is, return to the pre-event baseline) following the SPICE (Supplementary Fig. 9), which is unreasonable in light of the available constraints on SPICE duration[5]. Our simulations suggest, therefore, that the concentration of seawater sulphate was very low—at or below the low end of the 2–12 mM range suggested by previous work[14,15].

Another important result from the model is that the predicted ratio of carbon to sulphur (C/S) linked to this transient burial was very low: 1 to 4 moles of C per mole of S or 0.4–1.5 grams of C per gram of S (Fig. 4). In younger sediments, similar C/S ratios are observed only in sediments deposited under euxinic conditions[16]. The large scale of this Cambrian euxinia is suggested by comparison to the Black Sea, the largest modern euxinic basin. Our estimates for the transient burial flux of sulphur that caused the isotope excursion are equal to 50–75 times that of the euxinic portion of the Black Sea[17], thus providing the first quantitative evidence for widespread euxinia in the Palaeozoic ocean. Importantly, these estimates do not come close to predicting whole-ocean euxinia and probably reflect pervasive sulphidic mid-water-column conditions along basin margins.

Our argument for increased euxinia becomes stronger when we consider that ΔS may have decreased over the event (Fig. 2). ΔS (now preserved as $\Delta\delta^{34}S_{CAS-pyrite}$) is the isotopic offset between coexisting seawater sulphate (preserved as CAS) and dissolved hydrogen sulphide (preserved as pyrite) that results from microbial sulphate reduction and associated microbial pathways that lead to pyrite formation. For the two sections that have sufficient pyrite for isotopic analysis (eastern Laurentia and Gondwana), there is a strikingly systematic negative shift for ΔS parallel to the positive excursions for $\delta^{34}S_{CAS}$ and $\delta^{13}C_{carb}$ (Fig. 2). This shift is important because a smaller ΔS, when applied to our model, requires even greater pyrite burial to explain the positive sulphur excursion. The further increase in pyrite burial results in a lower mean C/S ratio, strengthening the case for burial under euxinic conditions (see Supplementary Fig. 12 for sensitivity tests of ΔS). Our Cambrian data must record a decrease in seawater sulphate concentration associated with progressive, voluminous euxinic pyrite burial during the SPICE under generally low levels of sulphate.

Additional evidence for the expansion of euxinic conditions comes from the coeval Alum Shale in Sweden, where a systematic decrease in Mo enrichment coincides with the SPICE (Fig. 3). Molybdenum is a transition metal typically enriched in organic-rich sediments deposited under euxinic conditions, with extents of enrichment tracking overall Mo availability in the local or global system as well as concentrations of organic carbon[18,19]. The decreasing enrichment seen in the Alum Shale is also apparent when Mo concentrations are normalized to total organic carbon content, suggesting that organic content is not the primary control. Importantly, the variability in Mo concentrations and Mo/TOC ratios occur despite iron proxy data that point to persistent euxinia over the interval of interest (Fig. 3); the Alum basin appears to have been locally euxinic before, during and after the SPICE.

The suggestion then is that another process drove the scale of enrichment. In short, the decline going into the SPICE and increase coming out argue for a decrease in the global Mo inventory of seawater as euxinic conditions expanded and then contracted on a globally significant scale[8,19,20]—a scenario consistent with the predictions of the modelled C and S data. We envision conditions during the SPICE to have been analogous to those during ocean anoxic events of the Mesozoic era, when the spread of euxinia led to extensive deposition of organic-rich, pyritic sediments in the deeper ocean, yielding concomitant isotopic shifts in dissolved inorganic carbon[21] and seawater sulphate[22].

The geochemical and stratigraphic framework of the SPICE provides new insight into the pronounced biological turnover associated with this event. Taken together with evidence for an initial sea-level rise, the geochemical data suggest that shoaling of toxic anoxic waters onto the shelf led to the extinction of shelf fauna, a situation similar to that envisioned for end-Permian extinctions[23]. Such a scenario was proposed previously to explain recurrent later Cambrian trilobite extinctions[12] but in acknowledged absence of independent constraints for such conditions.

Additional oscillations observed in the later Cambrian marine $\delta^{13}C$ record could reflect environmental perturbations similar to those linked to the SPICE. We suggest that anoxic water masses occurred widely in the subsurface of the later Cambrian ocean (that is, below the wind-mixed surface layer), a view that finds qualitative support in the stratigraphic distribution of organic-rich, pyritic black shales that peak in abundance in later Cambrian successions[24]. If correct, the high rates of biological turnover[25] and repeated trilobite extinctions[12,26] documented for later Cambrian fossils can be at least partially explained by episodic expansion of oxygen-depleted waters. In larger terms, broad patterns of Cambrian animal evolution may reflect persistent oxygen deficiency in subsurface waters of Cambrian oceans, shedding new light on early evolution of the Phanerozoic biosphere in the wake of late Proterozoic oxygenation.

Figure 4 | Examples of the modelled carbon and sulphur isotope composition of the ocean during the SPICE. To generate the isotope excursions, we doubled organic carbon burial from 4.1×10^{18} to 8.2×10^{18} moles per million years and increased pyrite burial from the steady-state rate (0.98×10^{18} moles per million years) by the factors given for half a million years. Values in parentheses are the molar carbon-to-sulphur (C/S) ratio of the transient burial fluxes introduced into the model. The starting marine sulphate concentration in these simulations was 1.5 mM. DIC, dissolved inorganic carbon.

1. Hough, M. L. *et al.* A major sulphur isotope event at c. 510 Ma: a possible anoxia–extinction–volcanism connection during the Early–Middle Cambrian transition? *Terra Nova* **18**, 257–263 (2006).
2. Hurtgen, M. T., Pruss, S. B. & Knoll, A. H. Evaluating the relationship between the carbon and sulfur cycles in the later Cambrian ocean: an example from the Port au
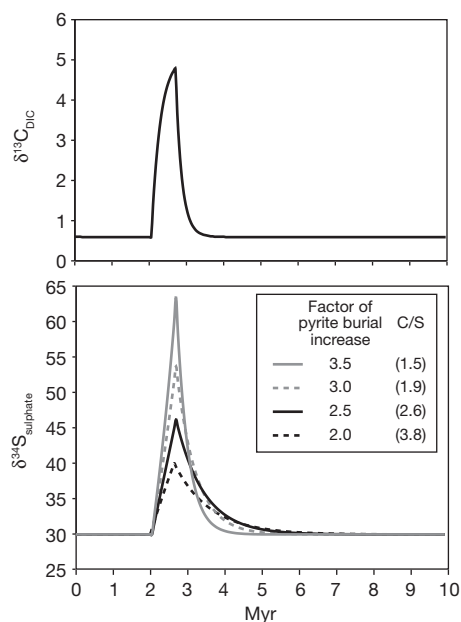
Port Group, western Newfoundland, Canada. *Earth Planet. Sci. Lett.* **281,** 288–297 (2009).

3. Zhuravlev, A. & Wood, R. Anoxia as the cause of the mid-Early Cambrian (Botomian) extinction event. *Geology* **24,** 311–314 (1996).

4. Saltzman, M. *et al.* A global carbon isotope excursion (SPICE) during the Late Cambrian: relation to trilobite extinctions, organic-matter burial and sea level. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **162,** 211–223 (2000).

5. Saltzman, M. R. *et al.* The Late Cambrian SPICE (δ^{13}C) Event and the Sauk II-SAUK III Regression: new evidence from Laurentian basins in Utah, Iowa and Newfoundland. *J. Sedim. Res.* **74,** 366–377 (2004).

6. Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396,** 450–453 (1998).

7. Poulton, S. W., Fralick, P. W. & Canfield, D. E. The transition to a sulphidic ocean ~1.84 billion years ago. *Nature* **431,** 173–177 (2004).

8. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452,** 456–459 (2008).

9. Brasier, M. D., Corfield, R. M., Derry, L. A., Rozanov, A. Y. & Zhuravlev, A. Y. Multiple δ^{13}C excursions spanning the Cambrian explosion to the Botomian crisis in Siberia. *Geology* **22,** 455–458 (1994).

10. Saltzman, M. R., Runnegar, B. & Lohmann, K. C. Carbon isotope stratigraphy of Upper Cambrian (Steptoean Stage) sequences of the eastern Great Basin: record of a global oceanographic event. *Geol. Soc. Am. Bull.* **110,** 285–297 (1998).

11. Montanez, I. P., Osleger, D. A., Banner, J. L., Mack, L. E. & Musgrove, M. Evolution of the Sr and C isotope composition of Cambrian Oceans. *GSA Today* **10,** 1–7 (2000).

12. Palmer, A. The biomere problem: evolution of an idea. *J. Paleontol.* **58,** 599–611 (1984).

13. Peng, S. *et al.* Global standard stratotype—section and point of the Furongian series and Paibian stage Cambrian. *Lethaia* **37,** 365–379 (2004).

14. Brennan, S. T., Lowenstein, T. K. & Horita, J. Seawater chemistry and the advent of biocalcification. *Geology* **32,** 473–476 (2004).

15. Gill, B. C., Lyons, T. W. & Saltzman, M. R. Parallel, high-resolution carbon and sulfur isotope records of the evolving Paleozoic marine sulfur reservoir. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **256,** 156–173 (2007).

16. Berner, R. Sedimentary pyrite formation: an update. *Geochim. Cosmochim. Acta* **48,** 605–615 (1984).

17. Neretin, L. N., Volkov, I. I., Böttcher, M. E. & Grinenko, V. A. A sulfur budget for the Black Sea anoxic zone. *Deep Sea Res. I* **48,** 2569–2593 (2001).

18. Emerson, S. & Huested, S. Ocean anoxia and the concentrations of molybdenum and vanadium in seawater. *Mar. Chem.* **34,** 177–196 (1991).

19. Algeo, T. J. & Lyons, T. W. Mo–total organic carbon covariation in modern anoxic marine environments: Implications for analysis of paleoredox and paleohydrographic conditions. *Paleoceanography* **21,** PA1016 (2006).

20. Algeo, T. J. Can marine anoxic events draw down the trace element inventory of seawater? *Geology* **32,** 1057–1060 (2004).

21. Arthur, M. A., Dean, W. E. & Pratt, L. M. Geochemical and climatic effects of increased marine organic carbon burial at the Cenomanian/Turonian boundary. *Nature* **335,** 714–717 (1988).

22. Adams, D. D., Hurtgen, M. T. & Sageman, B. B. Volcanic triggering of a biogeochemical cascade during Oceanic Anoxic Event 2. *Nature Geosci.* **3,** 1–4 (2010).

23. Wignall, P. B. & Twitchett, R. J. Oceanic anoxia and the end Permian mass extinction. *Science* **272,** 1155–1158 (1996).

24. Berry, W. B. N. & Wilde, P. Progressive ventilation of the oceans; an explanation for the distribution of the lower Paleozoic black shales. *Am. J. Sci.* **278,** 257–275 (1978).

25. Bambach, R. K., Knoll, A. H. & Wang, S. C. Origination, extinction, and mass depletions of marine diversity. *Paleobiology* **30,** 522–542 (2004).

26. Palmer, A. R. Biomere: a new kind of biostratigraphic unit. *J. Paleontol.* **39,** 149–153 (1965).

27. Scotese, C. R. *Atlas of Earth History* (PALEOMAP Project, 2001).

28. Ahlberg, P. *et al.* Cambrian high-resolution biostratigraphy and carbon isotope chemostratigraphy in Scania, Sweden: first record of the SPICE and DICE excursions in Scandinavia. *Lethaia* **42,** 2–16 (2008).

29. Lyons, T. W. & Severmann, S. A critical look at iron paleoredox proxies: new insights from modern euxinic marine basins. *Geochim. Cosmochim. Acta* **70,** 5698–5722 (2006).

30. Raiswell, R., Buckley, F., Berner, R. A. & Anderson, T. F. Degree of pyritization of iron as a paleoenvironmental indicator of bottom-water oxygenation. *J. Sedim. Res.* **58,** 812–819 (1988).

LETTER

Competition and phylogeny determine community structure in Müllerian co-mimics

Markos A. Alexandrou[1], Claudio Oliveira[2], Marjorie Maillard[1], Rona A. R. McGill[3], Jason Newton[3], Simon Creer[1] & Martin I. Taylor[1]

Until recently, the study of negative and antagonistic interactions (for example, competition and predation) has dominated our understanding of community structure, maintenance and assembly[1]. Nevertheless, a recent theoretical model suggests that positive interactions (for example, mutualisms) may counterbalance competition, facilitating long-term coexistence even among ecologically undifferentiated species[2]. Müllerian mimics are mutualists that share the costs of predator education[3] and are therefore ideally suited for the investigation of positive and negative interactions in community dynamics. The sole empirical test of this model in a Müllerian mimetic community supports the prediction that positive interactions outweigh the negative effects of spatial overlap[4] (without quantifying resource acquisition). Understanding the role of trophic niche partitioning in facilitating the evolution and stability of Müllerian mimetic communities is now of critical importance, but has yet to be formally investigated. Here we show that resource partitioning and phylogeny determine community structure and outweigh the positive effects of Müllerian mimicry in a species-rich group of neotropical catfishes. From multiple, independent reproductively isolated allopatric communities displaying convergently evolved colour patterns, 92% consist of species that do not compete for resources. Significant differences in phylogenetically conserved traits (snout morphology and body size) were consistently linked to trait-specific resource acquisition. Thus, we report the first evidence, to our knowledge, that competition for trophic resources and phylogeny are pivotal factors in the stable evolution of Müllerian mimicry rings. More generally, our work demonstrates that competition for resources is likely to have a dominant role in the structuring of communities that are simultaneously subject to the effects of both positive and negative interactions.

Positive interactions, such as mutualistic associations, can have important roles in the maintenance of community structure, potentially outweighing the negative effects of competition for niche space[2]. The empirical evidence for such phenomena is biased towards plants[5,6], with a single study on mimetic butterfly communities[4] complementing such research. The study convincingly demonstrated that Müllerian mimic butterflies converge spatially, but it only considered habitat utilization, whereas resource consumption and trophic niche were indirectly inferred as likely correlates of other variables (such as forest structure, topography and flight height). However, until now the extent of trophic overlap or the evolution of morphological traits associated with resource acquisition has not been directly quantified in Müllerian mimetic communities and this limits our ability to infer the importance of mimicry in determining community structure. Combining ecological, phylogenetic and morphological analyses provides a valuable empirical test of the recent theoretical proposition[2] that positive interactions among competitors can promote multispecies coexistence and the consequences of these interactions in terms of species diversity[7]. Tropical freshwater fish offer a novel vertebrate perspective on the evolution of mimetic community structure, as they

inhabit discontinuous habitats and are exposed to a multitude of piscivorous predators[8].

The Corydoradinae (Teleostei; Siluriformes; Callichthyidae) are a species-rich group of freshwater catfishes that inhabit streams, rivers and floodplains throughout South America[9]. The genus *Corydoras* comprises the majority of the Corydoradinae, and is the most species-rich genus of catfish with over 150 described species and as many undescribed taxa[9,10]. The Corydoradinae are almost all benthic omnivorous detritivores, consuming algae, terrestrial and aquatic insects, annelids and zooplankton[11]. At many sites, as many as three almost identically coloured species coexist and aggregate in large mixed shoals[9,11,12], with each geographical location hosting a different shared colour pattern. These colour patterns include both cryptic and disruptive elements (for example, spots, counter-shading and eye bars) and putatively aposematic elements (for example, strongly contrasting black and white stripes, orange and black patches and conspicuously coloured spines). Interestingly, some colour patterns have also been adopted by species belonging to different families and orders (*Otocinclus*[13], *Brachyrhamdia* and *Serrapinnus*).

In most cases, coexisting Corydoradinae species differ in snout morphology and body size[11]. Recorded predators of the Corydoradinae include *Plagioscion squamosissimus* (Perciformes)[14] and *Hoplias malabaricus* (Characiformes)[11], whereas kingfishers, egrets and herons are the dominant avian predators of armoured catfishes[15]. Corydoradinae are protected by sharp, lockable pectoral and dorsal spines, tough scutes covering the side and dorsal surfaces[10], and toxins secreted from the axillary glands[16-18]. The widespread distribution of Corydoradinae, propensity to aggregate, shared colour patterns and post-capture defences make them a unique system to study the mechanisms underpinning community structure.

A taxonomically comprehensive (80% coverage) phylogeny was constructed from 425 taxa (Supplementary Table 1) using Bayesian and maximum likelihood methods to resolve the relationships within the Corydoradinae and determine whether colour patterns are the result of convergence or shared ancestry (Fig. 1). Differences in tree topologies were small between mitochondrial and nuclear data sets, and both identified nine major lineages (Supplementary Figs 1–4). Using the resulting phylogeny, we identified 52 species belonging to 24 different mimicry rings, composed of two or three species each (Fig. 2). All lineages included taxa that were members of mimicry rings with the exception of lineage 2. A comparison of topological positions of respective co-mimics shows that 92% of mimetic communities are composed of species belonging to evolutionarily distinct lineages. In the only two mimetic groups with species belonging to the same genetic lineage, these are not sister species and apparent convergence may be due to close genetic affinity rather than convergence. As all other shared patterns are the product of convergence in sympatry, and all Corydoradinae are well protected, we consider Müllerian mimicry[19] to be the most convincing explanation for the pattern convergence. Patterns signalling unprofitability to predators do not have to be

[1]Environment Centre Wales, Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, College of Natural Sciences, Bangor University, Bangor LL57 2UW, UK. [2]Departamento de Morfologia, Instituto de Biociencias, Universidade Estadual Paulista, 18618-970 Botucatu, SP, Brazil. [3]NERC Life Sciences Mass Spectrometry Facility, Scottish Universities Environmental Research Centre, Rankine Avenue, East Kilbride G75 0QF, UK.
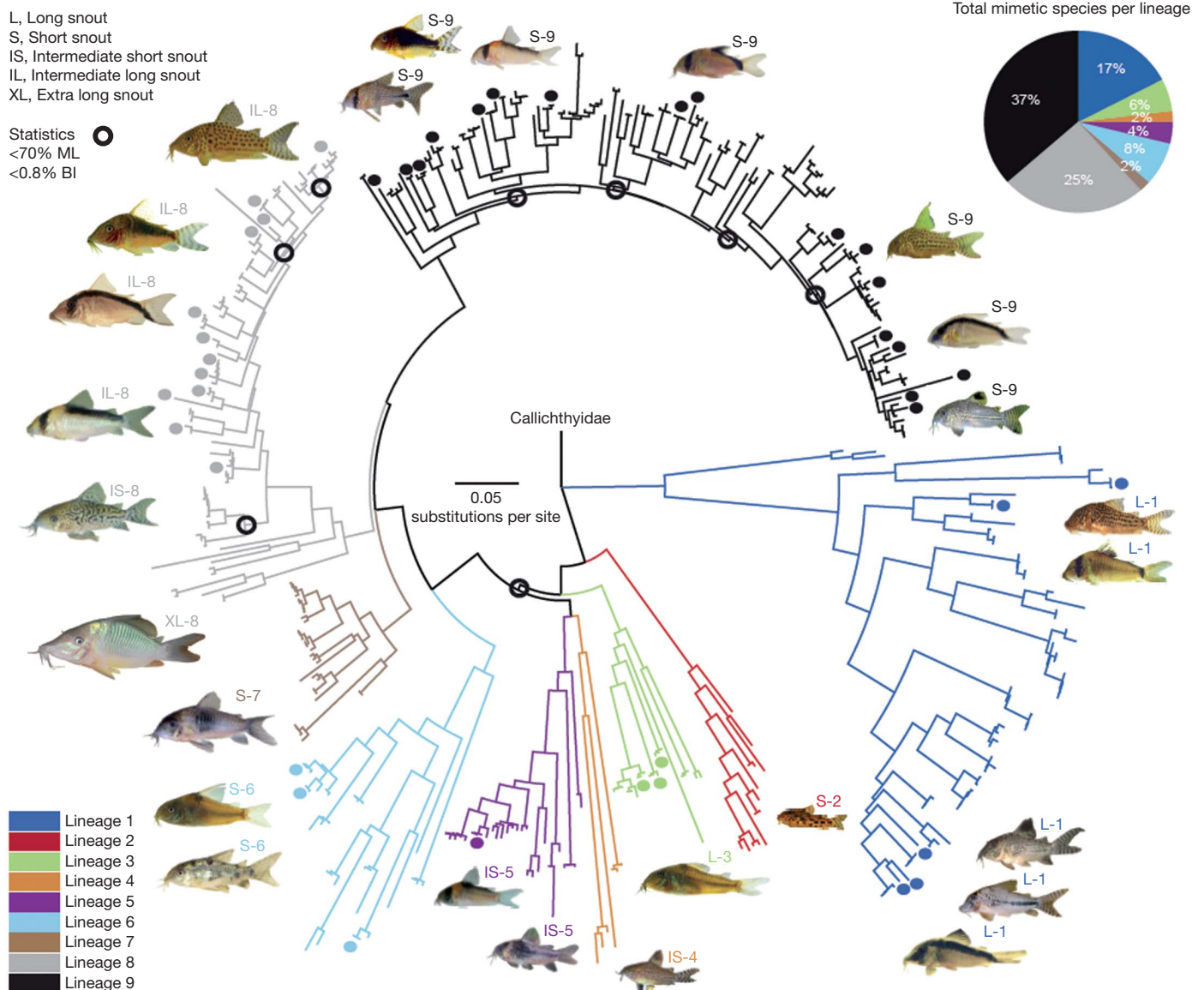
Figure 1 | Phylogenetic relationships of Corydoradinae including co-mimics. The pie chart shows percentage of mimetic species per lineage. Branches with mimetic species at tips are indicated with coloured circles (coded by lineage). Nodes with support below 0.8 (Bayesian inference; BI) probability and 70% (maximum likelihood; ML) are denoted with black open circles. Codes on pictures indicate snout types as determined by morphometrics and genetic lineage (L, long; S, short; IS, intermediate short; XL, extra long; IL, intermediate long). Representative images of morphotypes and colour patterns clockwise from lineage 1: (L-1) *Corydoras maculifer*, (L-1) *C. simulatus*, (L-1) *C.* sp. C109, (L-1) *C.* sp. C92, (L-1) *C. narcissus*; (S-2) *A. poecilius**; (L-3) *S. prionotus*; (IS-4) *C. mamore*; (IS-5) *C.* sp. CW19, (IS-5) *C. nijsseni*; (S-6) *C. paleatus*, (S-6) *C. nattereri*; (S-7) *C.* sp. CW26; (XL-8) *C. multiradiatus**; (IS-8) *C. sodalis**; (IL-8) *C. imitator*, (IL-8) *C.* sp. CW6, (IL-8) *C. seussi*, (IL-8) *C.* sp. C122; (S-9) *C.* sp. C91, (S-9) *C. gossei*, (S-9) *C. adolfoi*, (S-9) *C. metae*, (S-9) *C. araguaiaensis*, (S-9) *C. arcuatus*, (S-9) *C. julii*. *Non-mimetic taxa.

conspicuous but should be distinctive[20], while cryptic and aposematic elements present in individual prey are not necessarily mutually exclusive[21–23].

Of the 52 mimetic species, most belong to lineages 1, 8 and 9 (Fig. 1), suggesting a non-random frequency of co-occurrence between members of different lineages. In all cases, genetic distance is great enough between co-mimics for them to be considered reproductively isolated (mean pairwise mitochondrial distance = 11.16% ± 4.4 (standard deviation)). Furthermore, cytogenetic data indicate that respective members of different lineages have undergone extensive genomic duplications, with chromosome complements ranging from $2n = 44$–134 and genome sizes ranging from 1.1–8.75 pg[24,25]. Most of this variation occurs between lineages; however, there is also significant variation within lineages and often between sibling species.

Trophic interactions were elucidated using stable isotopes of carbon ($^{13}C/^{12}C$ ratio, reported as $\delta^{13}C$) and nitrogen ($^{15}N/^{14}N$ ratio, reported as $\delta^{15}N$), allowing us to assess the extent of dietary overlap among co-mimics. Nitrogen isotopes are particularly informative, as values increase in a stepwise manner between trophic levels; for example, carnivore tissues have higher $\delta^{15}N$ values than herbivore tissues[26]. Carbon isotope ratios ($\delta^{13}C$) may change slightly with trophic level, but the major source of variation has been attributed to differences in the sources of primary production, and $\delta^{13}C$ values are typically more useful in deriving foraging locations[27,28]. Significant differences ($P < 0.001$) after post-hoc corrections were found in mean $\delta^{15}N$ values within six mimicry rings composed of species belonging to different lineages, with different morphology (Table 1 and Supplementary Table 2). Niche overlap (no significant differences in mean $\delta^{15}N$)
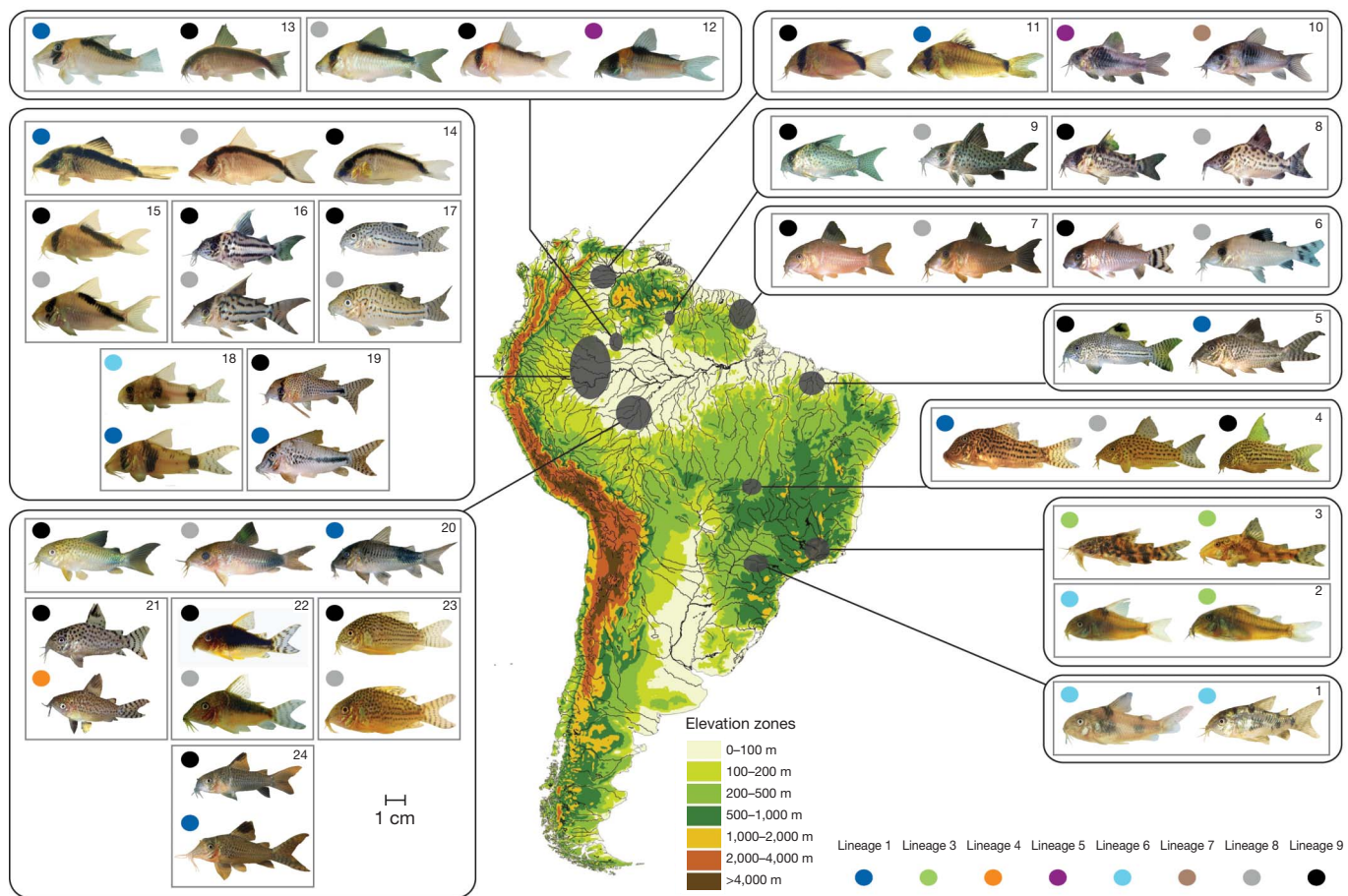
Figure 2 | Geographical distribution of mimetic communities. Genetic lineages are denoted by coloured circles; small grey rectangles represent independent mimetic communities numbered 1–24. Larger black rectangles indicate communities belonging to the same drainage or basin. Grey ellipses indicate approximate geographical distribution. Species images: (1) *C. paleatus*, *C. ehrhardti*; (2) *C. nattereri*, *Scleromystax prionotus*; (3) *S. barbatus*, *S. macropterus*; (4) *C. maculifer*, *C. sp. C122*, *C. araguaiaensis*; (5) *C. julii*, *C. sp. C109*; (6) *C. oiapoquensis*, *C. condisciplus*; (7) *C. sp. C135*, *C. sp. C136*; (8) *C.* *evelynae*, *C. sp. CW13*; (9) *C. kanei*, *C. crimmeni*; (10) *C. sp. CW19*, *C. sp. CW26*; (11) *C. metae*, *C. simulatus*; (12) *C. imitator*, *C. adolfoi*, *C. nijsseni*; (13) *C. serratus*, *C. cf. arcuatus*; (14) *C. narcissus*, *C. sp. CW6*, *C. arcuatus*; (15) *C. sp. C84*, *C. sp. C156*; (16) *C. sp. CW28*, *C. pulcher*; (17) *C. trilineatus*, *C. leopardus*; (18) *C. tukano*, *C. sp. CW11*; (19) *C. sp. C91*, *C. sp. C92*; (20) *C. similis*, *C. sp. C66*, *C. ourastigma*; (21) *C. cruziensis*, *C. mamore*; (22) *C. gossei*, *C. seussi*; (23) *C. sterbai*, *C. haraldshultzi*; (24) *C. sp. C76*, *C. sp. C77*.

was only observed between two co-mimic pairs (Supplementary Fig. 5). Differences in $\delta^{13}C$ values were significant between species in five out of the eight mimicry rings examined (Table 1 and Supplementary Table 2). Divergence of isotopic signatures reflects dietary segregation between long-snouted species compared to short-snouted species. There is a clear relationship between morphology and niche occupation in the Corydoradinae, as larger long-snouted species always occupy a lower relative 'trophic level' (consistently lower $\delta^{15}N$) than smaller

short-snouted species (Supplementary Table 2). Minor variation in isotopic signatures could also result from physiological differences between species. However, this is unlikely here given the scale of observed differences and the close genetic affinity of Corydoradinae.

Quantifying traits with adaptive value is valuable in the study of trophic differentiation. Body size and snout or jaw morphology are particularly important traits for food acquisition in fishes[29,30]. We used a landmark-based geometric morphometric approach, grouping our

Table 1 | Isotope statistics

Site	Species comparisons	$\delta^{13}C$ (P value)	$\delta^{15}N$ (P value)	$\delta^{13}C$ (d.f., F, P)	$\delta^{15}N$ (d.f., F, P)	Test type
4	*C. araguaiaensis* versus *C. maculifer*	0.606	0.00002	42, 0.679, 0.513	42, 52.588, 0.00000	One-way ANOVA plus Tukey–Kramer
4	*C. araguaiaensis* versus *C. sp. C122*	0.957	0.00126			
4	*C. maculifer* versus *C. sp. C122*	0.569	0.00029			
12	*C. adolfoi* versus *C. imitator*	0.00002	0.00002	41, 41.02, 0.00000	41, 46.104, 0.00000	One-way ANOVA plus Tukey–Kramer
12	*C. adolfoi* versus *C. nijsseni*	0.00002	0.00004			
12	*C. imitator* versus *C. nijsseni*	0.762	0.295			
13	*C. arcuatus* versus *C. serratus*	0.00000	0.00000	39, 47.062, 0.00000	39, 128.05, 0.00000	*t*-test plus Bonferroni
18	*C. tukano* versus *C. sp. CW11*	0.00000	0.00000	40, 27.878, 0.00000	40, 56.858, 0.00000	*t*-test plus Bonferroni
2	*C. nattereri* versus *S. prionotus*	0.00522	0.00000	35, 8.914, 0.00522	35, 43.837, 0.00000	*t*-test plus Bonferroni
3	*S. barbatus* versus *S. macropterus*	0.00000	0.186	41, 41.823, 0.00000	42, 1.808, 0.186	*t*-test plus Bonferroni
1	*C. paleatus* versus *C. ehrhardti*	0.275	0.842	28, 1.112, 0.275	28, 0.041, 0.842	*t*-test plus Bonferroni

Site 1, Rio Tibaji; site 2, Rio Fau; site 3, east coast; site 4, Rio Araguaia; site 12, Rio Negro; site 13, Rio Negro; site 18, Rio Tiquie. Statistical comparisons are shown between species pairs (P values) and mimicry rings within sites for carbon and nitrogen ratios (d.f., degrees of freedom; F, F test statistic; P, P value). Standard Bonferroni corrected critical P = 0.025.

results in principal component analysis (PCA) and canonical variate analysis (CVA) plots by phylogenetic lineage to assess the extent of morphological conservatism within and between lineages and co-mimics (Supplementary Figs 6 and 7). In 92% of cases examined, mimicry rings were found to be composed of species with significant morphometric differences, an observation confirmed with pairwise F tests comparing partial Procrustes distances between respective co-mimic lineages ($P < 0.01$ for all pairwise comparisons) (Supplementary Table 4). Key characters determining these differences include snout length, eye position and body depth, and these are probably important phenotypic traits facilitating niche differentiation, as they are directly associated with observed differences in niche occupation. Only two mimicry rings consist of species that are members of overlapping morphometric groups. Furthermore, observed morphological differences in snout and body size are phylogenetically conserved, indicative of a role for niche conservatism within lineages. Evidence from stable isotopes for eight mimicry rings and morphological analysis of all known communities suggests that ecologically relevant morphological divergence has resulted in snout-specific species assortment within mimetic communities more often than expected at random. Moreover, the majority of observed mimicry rings consist of species that have diverged in terms of resource acquisition, but converged in spatial occupation and colour pattern.

To investigate further the evolution of colour patterns among co-mimics we quantified the geographical distribution of colour patterns and compared them within and between communities. Mimetic Corydoradinae catfishes display a variety of contrasting colour pattern characteristics such as blocks of colour, bright spines, patches, bands, stripes, spots and reticulations. Using 20 different sections of the fish we scored the presence (1), absence (0), or variability (0.5) of different colour pattern characteristics (Methods). This allowed a quantification of pattern for each mimetic species that could then be compared to all other respective mimics using Euclidean pairwise distance matrices (Supplementary Fig. 8). A Mantel test revealed that sympatric co-mimics are more similar in colouration than those in allopatry ($r = -0.585$, $P = 0.001$), indicating a highly significant relationship between colour pattern and geographical distribution.

We report a diverse assemblage of vertebrate Müllerian mimics that inhabit aquatic environments through which they signal to predators foraging in two different optical modalities: water and air. We demonstrate that dietary resource partitioning coupled with morphological and phylogenetic differences determine community assembly despite the positive benefits of Müllerian mimicry. These results suggest that the benefits accrued by Müllerian co-mimics are not sufficient to overcome the need for ecological differentiation for stable long-term coexistence, thereby reinforcing the position that antagonistic interactions set ecological limits on the diversification of mutualistic communities. Despite these limits, Müllerian mimicry may increase diversification rates among allopatric communities if predation-driven directional selection simultaneously leads to convergence of colouration among sympatric taxa, but divergence among allopatric taxa. Many neotropical habitats are critically threatened by anthropogenic pollution, deforestation and river obstruction and modification. As a result we may lose many of these unique species and communities before we fully appreciate their extraordinary diversity.

METHODS SUMMARY

A total of 425 taxa (226 operational taxonomic units) were collected or obtained from wild populations. Partial mitochondrial and nuclear DNA sequences were amplified, sequenced and aligned to construct a comprehensive molecular phylogeny for the Corydoradinae. Maximum likelihood and Bayesian analysis were performed on separate and concatenated data sets to test for phylogenetic relatedness of mimetic species. A total of 18 species belonging to eight different independent communities were collected to examine trophic interactions within mimicry rings using stable isotopes of carbon and nitrogen. Individual tissue samples were processed and analysed using continuous-flow isotope ratio mass spectrometry. Mimetic and non-mimetic species were photographed and

analysed using a geometric morphometric approach based on 24 homologous landmarks to identify ecologically relevant differences in shape. Multivariate plots of transformed morphometric distances were used to assess the extent of overlap in overall shape within and between different lineages of mimetic species. Colour patterns of mimetic species were quantified by dividing the lateral section of each fish into 20 subsections, which were then scored for the presence, absence, or variability of particular patterns. Using these data a matrix was created describing colour pattern based on 100 colour characters for 52 mimetic species, while a second matrix was created scoring species in terms of geographical distance (sympatric or allopatric). The two matrices were compared using a Mantel test to examine the relationship between colour patterns and geographical distance.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. May, R. M. The role of theory in ecology. *Am. Zool.* **21,** 903–910 (1981).
2. Gross, K. Positive interactions among competitors can produce species-rich communities. *Ecol. Lett.* **11,** 929–936 (2008).
3. Rowland, H. M., Ihalainen, E., Lindström, L., Mappes, J. & Speed, M. P. Co-mimics have a mutualistic relationship despite unequal defences. *Nature* **448,** 64–67 (2007).
4. Elias, M., Gompert, Z., Jiggins, C. & Willmott, K. Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. *PLoS Biol.* **6** (2008).
5. Brooker, R. W. *et al.* Facilitation in plant communities: the past, the present, and the future. *J. Ecol.* **96,** 18–34 (2008).
6. Callaway, R. M. Positive interactions among plants. *Bot. Rev.* **61,** 306–349 (1995).
7. Stachowicz, J. J. Mutualism, facilitation, and the structure of ecological communities. *Bioscience* **51,** 235–246 (2001).
8. Ruxton, G. D., Sherratt, T. N. & Speed, M. P. *Avoiding Attack* (Oxford Univ. Press, 2004).
9. Fuller, I. A. & Evers, H.-G. *Identifying Corydoradinae Catfish* (Ian Fuller Enterprises, 2005).
10. Nelson, J. *Fishes of the World* (John Wiley & Sons, 2006).
11. Nijssen, H. Revision of the Surinam catfishes of the genus *Corydoras* (Pisces; Siluriformes; Callichthyidae). *Beaufortia* **18,** 1–75 (1970).
12. Sands, D. D. *The Behaviour and Evolutionary Ecology of* Corydoras adolfoi *and* Corydoras imitator*: Studies on Two Species of Sympatric Catfish from the Upper Rio Negro, Brazil.* PhD thesis, Univ. Liverpool (1994).
13. Axenrot, T. E. & Kullander, S. O. *Corydoras diphyes* (Siluriformes: Callichthyidae) and *Otocinclus mimulus* (Siluriformes: Loricariidae), two new species of catfishes from Paraguay, a case of mimetic association. *Ichthyol. Explor. Freshwat.* **14,** 249–272 (2003).
14. Luz-Agostinho, K. D. G., Agostinho, A. A., Gomes, L. C. & Julio, H. F. Influence of flood pulses on diet composition and trophic relationships among piscivorous fish in the upper Parana River floodplain. *Hydrobiologia* **607,** 187–198 (2008).
15. Power, M. E. Depth distributions of armored catfish—Predator-induced resource avoidance. *Ecology* **65,** 523–528 (1984).
16. Greven, H., Flasbeck, T. & Passia, D. Axillary glands in the armoured catfish *Corydoras aeneus* (Callichthyidae, Siluriformes). *Verh. Ges. Ichthyol.* **6,** 65–69 (2006).
17. Kiehl, E., Rieger, C. & Greven, H. Axillary gland secretions contribute to the stress-induced discharge of a bactericidal substance in *Corydoras sterbai* (Callichthyidae, Siluriformes). *Verh. Ges. Ichthyol.* **6,** 111–115 (2006).
18. Wright, J. J. Diversity, phylogenetic distribution, and origins of venomous catfishes. *BMC Evol. Biol.* **9,** 282 (2009).
19. Müller, F. Über die vortheile der mimicry bei schmetterlingen. *Zool. Anz.* **1,** 54–55 (1878).
20. Endler, J. A. & Mappes, J. Predator mixes and the conspicuousness of aposematic signals. *Am. Nat.* **163,** 532–547 (2004).
21. Merilaita, S. & Ruxton, G. D. Aposematic signals and the relationship between conspicuousness and distinctiveness. *J. Theor. Biol.* **245,** 268–277 (2007).
22. Wüster, W. *et al.* Do aposematism and Batesian mimicry require bright colours? A test, using European viper markings. *Proc. R. Soc. Lond. B* **271,** 2495–2499 (2004).
23. Wright, J. J. Conservative coevolution of Müllerian mimicry in a group of rift lake catfish. *Evolution* doi:10.1111/j.1558-5646.2010.01149.x (22 October 2010).
24. Hinegardner, R. & Rosen, D. E. Cellular DNA content and evolution of Teleostean fishes. *Am. Nat.* **106,** 621–644 (1972).
25. Oliveira, C., Almeida-Toledo, L. F., Mori, L. & Toledo-Filho, S. A. Extensive chromosomal rearrangements and nuclear-DNA content changes in the evolution of the armored catfishes genus *Corydoras* (Pisces, Siluriformes, Callichthyidae). *J. Fish Biol.* **40,** 419–431 (1992).
26. Peterson, B. J. & Fry, B. Stable isotopes in ecosystem studies. *Annu. Rev. Ecol. Syst.* **18,** 293–320 (1987).
27. Rubenstein, D. R. & Hobson, K. A. From birds to butterflies: animal movement patterns and stable isotopes. *Trends Ecol. Evol.* **19,** 256–263 (2004).
28. West, J. B., Bowen, G. J., Cerling, T. E. & Ehleringer, J. R. Stable isotopes as one of nature's ecological recorders. *Trends Ecol. Evol.* **21,** 408–414 (2006).

29. Clabaut, C., Bunje, P. M. E., Salzburger, W. & Meyer, A. Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution* **61,** 560–578 (2007).

30. Genner, M. J., Turner, G. F., Barker, S. & Hawkins, S. J. Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecol. Lett.* **2,** 185–190 (1999).

Author Contributions M.I.T. conceived the study, contributed to all data collection, analysis and writing and supervised M.A.A.; M.A.A. conducted fieldwork, DNA sequencing, stable isotope analysis, morphology and colour pattern analysis, data analysis and writing. C.O. co-supervised M.A.A. and organized and participated in field sampling and writing. M.M. and R.A.R.M. conducted DNA sequencing and stable isotope analysis, respectively; J.N. provided stable isotope advice and guidance; and S.C. co-supervised M.A.A. and contributed to writing.

Author Information Sequence data have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) with accession numbers detailed in Supplementary Table 1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.I.T. (m.taylor@bangor.ac.uk).

METHODS

Sample acquisition and phylogenetic analysis. A total of 425 taxa (226 operational taxonomic units with multiple representatives when available) were obtained from wild populations collected by the authors, or purchased as wild-caught aquarium imports (Supplementary Table 1). Species for which we lack genetic material (*C. mamore, C. evelynae, C. ourastigma, C. crimmeni, C.* sp. CW19, *C.* sp. CW26, *C.* sp. C135, *C.* sp. C76, *C.* sp. C77) were identified and assigned to lineages on the basis of geometrical morphometric analysis, with which it is possible to differentiate between lineages. Partial sequences of 12S rRNA, 16S rRNA, ND4, tRNA[His], tRNA[Ser], cytochrome *b* (Cytb) and recombination activating gene (RAG1) were amplified for 425 taxa, and a nuclear intron from F-reticulon-4 amplified for 24 taxa, using the polymerase chain reaction (PCR) with the primers detailed in Supplementary Table 4. All products were sequenced in both directions using Big Dye terminator technology (Applied Biosystems). The quality of chromatograms was visually inspected and contigs were assembled using Geneious v 4.7 (ref. 31). A total of 50 alternative alignments for the 12s, 16s and F-reticulon-4 markers were generated using ProAlign[32], discarding unstable positions that differed more than 50% (gap opening penalty 7–15; gap extension penalty 3–7). The ND4, Cytb and RAG1 genes were aligned with MUSCLE[33], and all alignments were checked by eye. Substitution saturation and base compositional biases were tested for each gene and partition. Model selection was performed using the hLRT criterion under a fixed BIONJ-JC topology in JModelTest[34]. All data were partitioned by gene and coding position where appropriate. Incongruent length difference tests were performed to test for heterogeneity between nuclear and mitochondrial data sets. Subsequent analyses were performed on the following data sets: (1) MIT: all mitochondrial markers (for practical purposes, 12s + 16s considered as a single partition, as were tRNA[His] + tRNA[Ser]); (2) MITNUC: the MIT and RAG1 data combined; (3) MITNUC2: a smaller subset of the MITNUC data combined with the F-reticulon-4 intron for 25 representative taxa; (4) RAG1: a single nuclear marker independent of mitochondrial data. We rely primarily on the MIT data set as it represents the largest sample of linked markers, and use the others for comparisons. RAxML[35] using the web server RAxML BlackBox[36] was used for maximum likelihood analyses under a mixed partition model for all analyses. Random starting trees were used for each independent maximum likelihood (ML) tree search and all other parameters were set on default. Topological robustness was investigated using 500 non-parametric bootstrap replicates. Analyses were conducted under both GTR+G and GTR+I+G to assess whether implementing P-Invar and Gamma together affected parameter estimation. Bayesian analyses were performed using MrBayes v3.1.2 (ref. 37) Metropolis-coupled Markov chain Monte Carlo (MCMC) runs were set with random starting trees, one cold and three heated chains for 30 million generations, sampled every 1,000 generations for all data sets. We used the default uniform Dirichlet distribution for the base frequencies, and default prior distributions for all other parameters. Bayesian posterior probabilities were then calculated from the sample points after MCMC convergence. To ensure that analyses were not trapped in local optima, two independent MCMC runs were performed. Topologies and posterior clade probabilities from different runs were compared for congruence using Tracer 1.4 (ref. 38), to ensure adequate estimated sample sizes and ensure adequate mixing of parameters. All trees estimated before convergence were discarded. Trees from different runs were then combined using LogCombiner[39], and maximum clade credibility trees (mean node heights) were estimated using TreeAnnotator[39].

Geometric morphometrics. A total of 200 preserved individuals, representing over 120 different species (including all mimetic taxa), were photographed and used for digital landmark-based morphometric analysis of body shape (Supplementary Table 1; asterisks denote species analysed using morphometrics). We defined 24 easily identifiable homologous landmarks (see Supplementary Fig. 9), including two for scaling (ruler 1 cm). Photographs with landmarks were digitalized (using tpsDig by F. J. Rohlf, available at http://life.bio.sunysb.edu/morph/), and converted to coordinates under Procrustes superimposition (using CoordGen6 by H. D. Sheets, part of the Integrated Morphometrics Package (IMP) at http://www2.canisius.edu/~sheets/morphsoft.html) to describe shape change independent of size (removing potential ontogenetic effects and issues of allometry). We used PCAgen and CVAgen packages from IMP to create multivariate plots of Procrustes superimpositions (with axes normalized to lengths of one for all plots and outputs) and vectors of CV coefficients (scaled to a length of one) respectively. We also conducted pairwise comparisons of co-existing lineages to check for statistically significant differences in shape using Goodall's *F*-tests as implemented

in TwoGroup6 (Supplementary Table 3). Deformation plots were generated to identify landmarks contributing to shape changes across lineages and between co-mimics.

Stable isotope analysis. Muscle tissue was collected from 20 individuals (where possible) from 18 species belonging to eight different independent mimicry rings. All samples were dried overnight in an incubator at 60 °C to constant weight before storage in 1.5-ml centrifuge tubes containing silica gel. On return to the laboratory, the tissue samples were homogenized to a fine powder using a mortar and pestle. Approximately 0.7 mg of homogenized muscle was weighed out and distributed into tin capsules for analysis by continuous-flow isotope ratio mass spectrometry, using a Costech (model ECS 4010) elemental analyser with a ThermoFinnigan Delta Plus XP mass spectrometer. We used a number of gelatine and alanine standards (two for every ten samples) to obtain a standard deviation of 0.2‰ for $\delta^{15}N$ and 0.1‰ for $\delta^{13}C$. Samples of ground tryptophan were incorporated into each run as independent isotope standards and to calculate C and N abundance. These internal laboratory standards have been measured against secondary international isotope standards provided by the NIST and IAEA. All data are referenced to the primary international standards atmospheric nitrogen (AIR $\delta^{15}N$) and Vienna- PeeDee Belemnite (V-PDB $\delta^{13}C$). The effects of lipid extraction were checked on a subset of samples using a 10:5:4 methanol/chloroform/water extraction, repeated three times to obtain a clear supernatant[40]. Isotope values from lipid and non-lipid extracted samples of the same aliquot were the same. Separate aliquots of muscle tissue were used to obtain carbon and nitrogen isotope analyses. All resulting carbon and nitrogen stable isotope ratios were checked for conformity to a normal distribution and analysed with one-way ANOVAs for mimicry rings consisting of >2 species, followed by Tukey–Kramer post-hoc tests. In cases with only two species, we used a two-sample *t*-test to compare means, followed by Bonferroni post-hoc corrections. Significant differences in stable isotope ratios were then used to determine the degree of dietary differentiation to identify communities composed of species that partition trophic resources from those that do not. However, stomach contents analyses were not conducted and therefore these results cannot be extended to quantify dietary overlap (that is, the proportion of food items shared between coexisting species).

Colour pattern analysis. Pictures of live fishes were obtained from specimens caught in the wild and later kept under standardized aquarium conditions. Colour patterns (not hue) of mimetic species were quantified by dividing the lateral section of each fish into 20 subsections (Supplementary Fig. 9). Subsections were then scored for the presence (1), absence (0), or variability (0.5) of particular patterns (contrasting bands, reticulations, spots, blotches, bright patches, dark patches, stripes, uniform brown and lack of patterns). Using these data we created a matrix describing colour pattern based on 100 characters for 52 mimetic species, and generated a pairwise Euclidian distance matrix using MVSP 3.1 (ref. 41). A second pairwise matrix was created to score the respective members of independent mimicry rings in terms of geographical distribution, where sympatric species were scored 1, whereas allopatric species were scored 0. The significance of the relationship between the two matrices was investigated using a Mantel test with 10,000 permutations using the software zt[42].

31. Drummond, A. J. *et al.* Geneious v4.7 ⟨http://www.geneious.com⟩ (2009).
32. Loytynoja, A. & Milinkovitch, M. C. SOAP, cleaning multiple alignments from unstable blocks. *Bioinformatics* **17,** 573–574 (2001).
33. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bionf.* **5,** 1–19 (2004).
34. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25,** 1253–1256 (2008).
35. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
36. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57,** 758–771 (2008).
37. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).
38. Rambaut, A. & Drummond, A. J. Tracer v1.4 ⟨http://beast.bio.ed.ac.uk/Tracer⟩ (2004).
39. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7** (2007).
40. Pinnegar, J. K. & Polunin, N. V. C. Differential fractionation of $\delta^{13}C$ and $\delta^{15}N$ among fish tissues: implications for the study of trophic interactions. *Funct. Ecol.* **13,** 225–231 (1999).
41. Kovach, W. L. MVSP—A MultiVariate Statistical Package for Windows, version 3.1 (Kovach Computing Services, 1999).
42. Bonnet, E. & Van de Peer, Y. zt: a software tool for simple and partial Mantel tests. *J. Stat. Softw.* **7,** 1–12 (2002).

LETTER

Experimental niche evolution alters the strength of the diversity–productivity relationship

Dominique Gravel[1,2], Thomas Bell[3], Claire Barbera[2], Thierry Bouvier[4], Thomas Pommier[4,5], Patrick Venail[2]† & Nicolas Mouquet[2]

The relationship between biodiversity and ecosystem functioning (BEF) has become a cornerstone of community and ecosystem ecology[1–3] and an essential criterion for making decisions in conservation biology and policy planning[4,5]. It has recently been proposed that evolutionary history should influence the BEF relationship because it determines species traits and, thus, species' ability to exploit resources[6,7]. Here we test this hypothesis by combining experimental evolution with a BEF experiment. We isolated 20 bacterial strains from a marine environment and evolved each to be generalists or specialists[8]. We then tested the effect of evolutionary history on the strength of the BEF relationship with assemblages of 1 to 20 species constructed from the specialists, generalists and ancestors[9]. Assemblages of generalists were more productive on average because of their superior ability to exploit the environmental heterogeneity[10]. The slope of the BEF relationship was, however, stronger for the specialist assemblages because of enhanced niche complementarity. These results show how the BEF relationship depends critically on the legacy of past evolutionary events.

Two fundamental ecological mechanisms can generate positive BEF relationships[11,12]. First, species may occupy complementary ecological niches, for example by feeding on different resources. In communities of complementary species, more of the total available niche space is filled in diverse communities, resulting in better community-wide resource use. Second, high-functioning and competitively dominant species are more likely to be found within species-rich communities (the sampling effect). Both mechanisms require a detailed understanding of species' phenotypic traits[13,14]. There has been, however, virtually no effort to understand how the evolution of species traits within ecological communities affects ecosystem functioning[6,7,15].

One important trait that determines complementarity is the degree of resource specialization, that is, the number of resources a species is able to exploit. Species niche width will tend to evolve to match the amount of available environmental variation[16,17]. In simple environments specialized types are expected to evolve, whereas generalists are more likely to appear in environments containing many resources[18,19]. The degree of specialization could alter the BEF relationship[10], so a full understanding the relationship must account for the evolutionary forces driving trait diversity.

In communities containing only specialist species that feed on different resources, the species do not compete with each other and their effects on ecosystem functioning (here productivity) are therefore additive (that is, the BEF relationship is linear; Fig. 1). However, the increased ability to exploit one resource might come with a lower ability to exploit any other, that is, there is a trade-off between resource usage ability[18] (Fig. 1a). For any type of trade-off, evolution towards generalization will affect the BEF relationship (Fig. 1b and Supplementary Information, section 1). First, ecosystem functioning at low diversity should be lower for specialists because they are inefficient at exploiting environmental heterogeneity. Second, generalization should reduce the

contribution of additional species to ecosystem functioning and, thus, the slope (that is, strength) of the BEF relationship. Generalization also increases niche overlap and thus produces a nonlinear, saturating BEF relationship. We therefore predicted that the ecosystem functioning would be higher for generalists at low diversity and that the slope of the BEF relationship would be reduced in communities of generalists.

We tested these predictions by experimentally evolving niche specialization and conducting BEF experiments. Briefly, we promoted the evolution of generalist and specialist strategies from 20 ancestral bacterial strains that had been isolated from a marine environment. Each strain was grown either on a single resource (one different carbon substrate for each strain) or on a mixture of 31 resources (with a total resource availability equal to that in the single-resource treatment). Bacteria were serially transferred to fresh medium every 48 h for 32 transfers, allowing evolutionary adaptations[20] (Supplementary Information, section 2). Bacteria grown on the mixed medium (hereafter called generalists) tended to have higher performance on a wide array of substrates in comparison with the bacteria evolved on the simple medium consisting of a single resource (hereafter called specialists; see below).

We conducted the BEF experiment on the mixed medium of 31 resources for both evolutionary schemes (generalist and specialist) and with the ancestral strains (hereafter referred as species). We used the productivity (bacterial metabolic activity) after 48 and 72 h as our measure of ecosystem functioning[15]. The selection period resulted in a substantially increased productivity at all levels of species richness for both specialists and generalist assemblages (Fig. 2). Productivity significantly increased with the logarithm of species richness (48 h, $F_{1,1,506} = 291.2$, $P < 0.001$; 72 h, $F_{1,1,506} = 179.2$, $P < 0.001$), indicating that productivity was a saturating function of species richness. Productivity in monocultures differed significantly among treatments (48 h, $F_{2,1,506} = 1,751.9$, $P < 0.001$; 72 h, $F_{2,1,506} = 2,309.3$, $P < 0.001$), with ancestors performing the worst, followed in order by specialists and generalists. The slope of the BEF relationship also differed significantly among treatments (48 h, $F_{2,1,506} = 16.2$, $P < 0.001$; 72 h, $F_{1,1,506} = 8.8$, $P < 0.001$), being steeper for specialists. A model accounting for species composition also showed that the BEF relationship of ancestors was best described by a linear function of species richness and a nonlinear (saturating) function for the two evolutionary treatments (Supplementary Information, section 3). There was no relationship between the contribution of ancestors to the BEF and the contribution of their evolved counterparts (Supplementary Information, section 4). The difference between the slopes of the specialist and generalist BEF relationships was even stronger after 72 h, a result similar to those of experiments conducted with plant[21,22] and marine[23] communities.

We investigated whether the difference in the strength of the BEF relationship resulted from specialization, by growing each of the ancestral strains and specialist and generalist lineages on the 31 individual carbon substrates to estimate their final niche width. We recorded the number

[1]Université du Québec à Rimouski, Département de Biologie, Chimie et Géographie, 300 Allée des Ursulines, Québec G5L 3A1, Canada. [2]Institut des Sciences de l'Evolution UMR 5554, Centre National de la Recherche Scientifique, Université Montpellier 2, CC 065, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France. [3]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [4]Ecosystèmes Lagunaires UMR 5119, Centre National de la Recherche Scientifique, Université Montpellier 2, CC 093, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France. [5]Laboratoire d'Ecologie Microbienne (UMR 5557, USC 1193), Université Lyon I, INRA, CNRS, Bâtiment G. Mendel, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France. †Present address: Centro de Investigaciones Microbiológicas, Universidad de los Andes, Carrera 1A No. 18A-10, Oficina A305, Bogotá, Colombia.
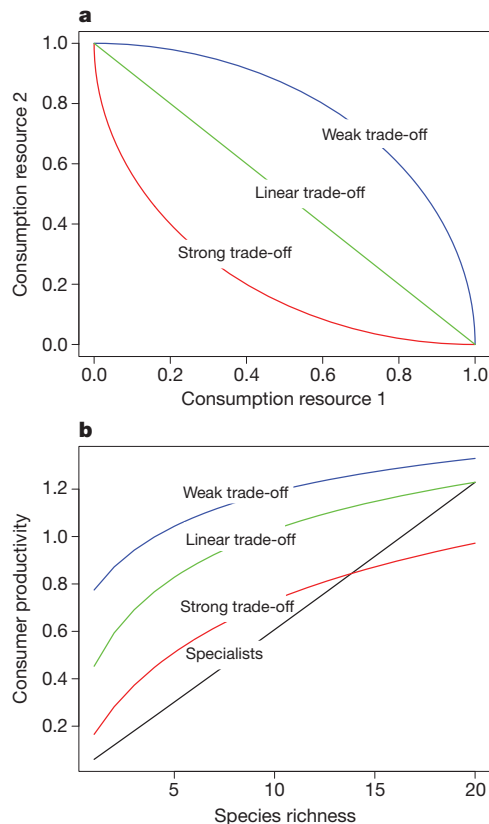
Figure 1 | Theoretical predictions of the effect of niche specialization on the strength of the BEF relationship. a, Hypothetical relationship between the consumption rates of a generalist on two resources. A perfect trade-off follows a straight line and any deviation to this reflects a relative cost (strong trade-off) or benefit (weak trade-off) of generalization. **b,** The BEF relationship will be affected by generalization and the type of trade-off. The figure represents a simulation experiment with $N = 20$ resources (R_i) and up to as many ($S \leq N$) consumer species (C_j). The dynamics of this system is given by the simple chemostat model $dR_i/dt = eR_0 - eR_i - \sum_{j=1}^{S} \alpha_{ij}R_iC_j$, $dC_j/dt = \sum_{i=1}^{N} \alpha_{ij}R_iC_j - mC_j - eC_j$, where e is the dilution rate, R_0 is the resource concentration in the inflow (for simplicity, we suppose that all resources have the same concentration in the inflow), α_{ij} is the per capita consumption rate of the resource i by consumer j, and m is the mortality rate. The ecosystem productivity at equilibrium is equivalent to the total resource consumption and is given by $\Phi = \sum_{i}^{N} \sum_{j}^{S} \alpha_{ij}R_iC_j$. Our simulation parameters are $R_0 = 1$, $e = 0.1$ and $m = 0.1$. We specified for the specialist that the consumption rate for its preferential resource is $\alpha_{ii} = 1$ and $\alpha_{i,j \neq i} = 0$ for the alternative resources. For the generalists, we specified the performance on the preferential resource to be $\alpha_{ii} = 0.6$ and an equal partitioning of the consumption rates between alternative resources that sums to 0.4 for the linear trade-off, 0.1 for the strong trade-off and 0.55 for the weak trade-off. We simulated communities of 1 to 20 species for specialists and generalists with linear, weak and strong trade-offs. Further details of the model and analytical results are given in Supplementary Information.

of substrates each lineage was able to exploit. Bacteria cultured on single substrates adapted to fewer substrates than bacteria cultured on mixed medium (Table 1 and Supplementary Information, section 5). The generalists were able on average to exploit a larger number of substrates (10.75 ± 1.49 (s.e.) of the 31 substrates) than the specialists (4.80 ± 0.51 (s.e.); t-test for paired samples, $t_{19} = 3.45$, $P = 0.002$). The average number of shared substrates between all pairs of strains and lineages was also much higher for generalists than for ancestors and specialists (Table 1). This ability to exploit more substrates is reflected in the performance of the bacteria when grown on the mixed medium. The maximal performance recorded for each substrate for generalists was also significantly higher than the maximal performance for specialists (t-test for paired samples, $t_{30} = 2.95$, $P = 0.006$; Table 1), which suggests there was no trade-off in resource usage ability. It has been
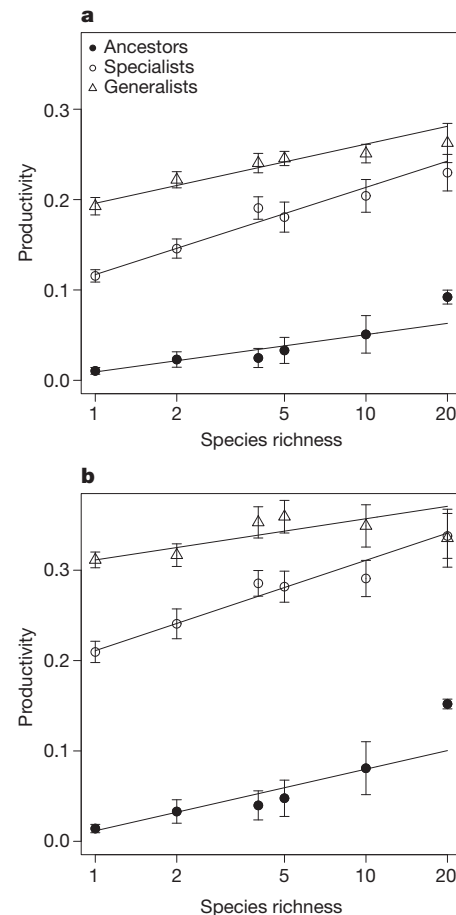

Figure 2 | Evolutionary treatments affect the strength of the biodiversity–productivity relationships. Productivity is measured as the absorbance at 590 nm after 48 h (corresponding to the conditions experienced during the selection experiment; **a**) and 72 h (**b**). Data show mean $\pm$ s.e. ($n = 504$ per evolutionary treatment). Lines depict the results of the analysis of covariance model.

shown that concurrent adaptation to multiple resources does not always limit the capability to exploit each resource individually[8]. In fact, mutations increasing fitness on a given resource can even sometimes increase fitness on other resources (that is, synclinal selection[24]), preventing the occurrence of trade-offs. We also checked for within-lineage genotypic variability and found it to be low (Supplementary Information, section 6). Most generalist lineages were composed of generalist genotypes, but some consisted of mixtures of coexisting specialist genotypes. However, the qualitative results and conclusions (Fig. 2) were unaffected when the analysis was restricted to only those lineages that were composed of the most generalist genotypes (Supplementary Information, section 6).

Some lineages that evolved on the mixed medium responded weakly to the experimental evolution and remained specialists, and some lineages from the specialist treatment evolved towards generalization (Supplementary Information, section 5). Consequently, there is variance in ecosystem functioning that should be better predicted by functional diversity[7]. In addition to enhanced complementarity, better performance of the generalists on each carbon source might also have contributed to the difference in the BEF relationships between evolutionary treatments (Fig. 2). We calculated a niche diversity index (NDI) for every assemblage of the specialist and generalist treatments. The NDI is the number of different substrates a community is able to exploit, calculated on the basis of assays of individual species[25] (Supplementary Information, section 7). Given that generalists were more productive on each substrate, we expected that for a given NDI

Table 1 | The effect of experimental evolution on specialization

	Ancestral strains	Specialist lineages	Generalist lineages
No. of substrates used	4.00 ± 0.69	4.80 ± 0.51	10.75 ± 1.49
No. of substrates used by the 20 species assemblages	16	14	29
Average no. of shared substrates between strain–lineage pairs	2.25 ± 0.08	2.74 ± 0.11	5.80 ± 0.24
Average maximal productivity for each substrate	0.21 ± 0.07	0.23 ± 0.08	0.42 ± 0.10
Productivity of monocultures on the mixed medium	0.01 ± 0.01	0.12 ± 0.01	0.19 ± 0.02

Niche specialization was assessed from performance assays on the 31 carbon sources. A strain (or lineage) was considered able to exploit a substrate when its absorbance at 590 nm after 48 h was larger than the 95% distribution of the blanks. All data, mean $\pm$ s.e.

productivity would be larger for the generalist assemblages. We found that productivity significantly increases with community NDI ($F_{1,667} = 84.0$, $P < 0.001$; Fig. 3) (we note that the analysis is conducted on the NDI range from 3 to 14 to meet the requirements of analysis of covariance) and the logarithm of species richness ($F_{1,667} = 5.23$, $P = 0.022$). The intercepts and the slope of the NDI–productivity relationship differ between evolutionary treatments (48 h, $F_{1,667} = 72.6$, $P < 0.001$; 72 h, $F_{1,667} = 19.4$, $P < 0.001$). Overall, most of the variance is accounted for by the NDI and the effect of the evolutionary treatments on the intercept. The larger amount of explained variance by the NDI argues for complementarity as the dominant mechanism. Our experimental evolutionary treatments therefore affected both species complementarity and maximal productivity at equivalent complementarity.

The ancestral strains had not previously encountered the experimental conditions, so it is unsurprising that the intercept of the BEF relationship was greatly reduced as a result of their maladaptation (Fig. 2). Nonetheless, there is still a significant positive BEF relationship. Because each species was equally represented in the experiment, it is possible to estimate the degree to which they were associated with higher- or lower-than-average levels of functioning[26]. We found that the inferred species contributions were dominated by a single ancestral strain, whereas they were distributed more equitably in the specialist and the generalist treatments (Supplementary Information, sections 3 and 4). We note there is no significant correlation between the ancestors' contribution to the BEF and the contribution of their evolved counterparts. The data therefore provide evidence that evolutionary history could affect both mechanisms of the BEF relationship. First, the BEF relationship will be stronger for communities of specialists because of enhanced complementarity. Second, if most species are maladapted, few species are able to contribute to functioning and

sampling effects dominate. Such a mechanism might be particularly important for ecosystem functioning in variable environments, where species are far from their optimal fitness peaks[27].

In this study, we have deliberately evolved independent lineages of specialists and generalists to compare assemblages of species that come from the same ancestral strain but have different evolutionary histories. It is likely that in nature species will diversify in complex assemblages of specialists and generalists depending on the environmental context[28]. We note that we cannot exclude the possibility that the generalists were an ensemble of specialist genotypes. Previous work has, however, shown that selection in a heterogeneous environment is most likely to result in the evolution of generalists[8]. In any case, the genotypic variability in a population, leading to increased generality, is still a species' trait that will influence the BEF relationship. Some of the changes observed between the evolved lineages might also have come through physiological adaptation. However, after at least a hundred generations we have found highly contrasting metabolic profiles (Table 1), no correlation between species contributions to the BEF relationship (Supplementary Information, section 3) and a strong response to selection (Supplementary Information, section 2). All of these observations are consistent with evolutionary changes.

A variety of BEF relationships have been observed and different ecological mechanisms have been inferred[22,29,30]. Our results provide strong support for the role of complementarity and evolutionary history in BEF. We found that specialists contribute more to the BEF. Monocultures of generalists were also found among the most productive assemblages. For conservation decisions, these results emphasize that, on average, the loss of specialists will have stronger effects on ecosystem functioning, but that losing a generalist species might have disproportionate effects when there is low redundancy. Our understanding of the mechanisms underlying the BEF relationship has now moved to a point where we cannot only distinguish among mechanisms, but can also manipulate these mechanisms experimentally. Investigations should now turn to understanding the evolutionary pressures that maintain niche diversification in natural communities, along with the trade-offs involved, and their effect on the BEF relationship.

METHODS SUMMARY

We isolated 31 phage-free bacterial strains from coastal sea water sampled off the Bay of Blanes, Spain, on the basis of their morphologies. We sequenced the 16S ribosomal DNA genes of the ancestral strains to confirm that different taxa were used in the experiment (Supplementary Information, section 8). Each strain individually underwent selection on a different, single-carbon substrate of an EcoPlate to obtain specialists, and underwent selection on a highly mixed medium made from a mixture of all 31 EcoPlate carbon substrates to obtain generalists. We transferred the bacteria to a fresh medium every 48 h during 64 d of incubation at 20 °C. After the selection period, we kept 20 lineages from among those that persisted and conducted a BEF experiment. We assembled microcosms with diversity levels, s, of 1, 2, 4, 5, 10 and 20 species for each of the three treatments (ancestral strains and specialist and generalist lineages). For each evolutionary treatment and diversity level, we created $20/s$ different assemblages by randomly selecting species from the species pool without replacement (for example, if $s = 5$ we randomly assigned the 20 species to four assemblages)[9], for a total of 42 assemblages. We carried out this process independently four times, so there were a total of 168 different assemblages for each evolutionary treatment. Each assemblage was replicated three times for a total of 1,512 microcosms (the product of three treatments, three replicates and 168 assemblages). We measured the light absorbance at 590 nm after 48 and 72 h to approximate productivity (reported
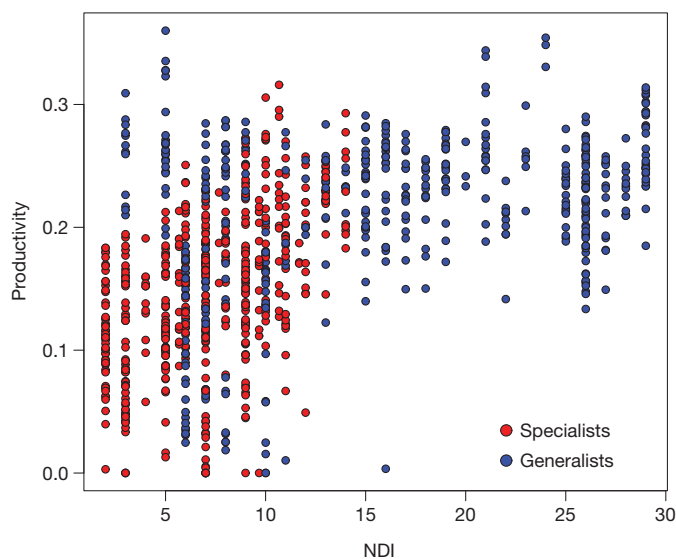

Figure 3 | The relationship between NDI and ecosystem functioning. NDI is the total number of carbon substrates a community is able to exploit, assessed from the individual ability of each lineage to exploit the carbon substrates and the community composition.

values were corrected by removing the average value of the blanks). We also conducted assays for each ancestral strain and the specialist and generalists lineages by incubating them on the 31 carbon substrates for 48 h at 20 °C. The assays were used to quantify generality and the niche diversity in the assemblages.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Kinzig, A. P., Pacala, S. & Tilman, G. D. (eds) *The Functional Consequences of Biodiversity: Empirical Progress and Theoretical Extensions* (Princeton Univ. Press, 2002).
2. Loreau, M., Naeem, S. & Inchausti, P. (eds) *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (Oxford Univ. Press, 2002).
3. Naeem, S., *et al. Biodiversity, Ecosystem Functioning, and Human Wellbeing: An Ecological and Economic Perspective* (Oxford Univ. Press, 2009).
4. Srivastava, D. S. & Vellend, M. Biodiversity-ecosystem function research: is it relevant to conservation? *Annu. Rev. Ecol. Evol. Syst.* **36,** 267–294 (2005).
5. Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ.* **7,** 437–444 (2009).
6. Cadotte, M. W., Cardinale, B. J. & Oakley, T. H. Evolutionary history and the effect of biodiversity on plant productivity. *Proc. Natl Acad. Sci. USA* **105,** 17012–17017 (2008).
7. Cadotte, M. W., Cavender-Bares, J., Tilman, D. & Oakley, T. H. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PLoS ONE* **4,** e5695 (2009).
8. Barrett, R. D. H., MacLean, R. C. & Bell, G. Experimental evolution of *Pseudomonas fluorescens* in simple and complex environments. *Am. Nat.* **166,** 470–480 (2005).
9. Bell, T. *et al.* The contribution of species richness and composition to bacterial services. *Nature* **436,** 1157–1160 (2005).
10. Loreau, M. Microbial diversity, producer-decomposer interactions and ecosystem processes: a theoretical model. *Proc. R. Soc. Lond. B* **268,** 303–309 (2001).
11. Tilman, D., Lehman, C. L. & Thomson, K. T. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94,** 1857–1861 (1997).
12. Loreau, M. Biodiversity and ecosystem functioning: a mechanistic model. *Proc. Natl Acad. Sci. USA* **95,** 5632–5636 (1998).
13. Diaz, S. & Cabido, M. Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* **16,** 646–655 (2001).
14. Reiss, J., Bridle, J. R., Montoya, J. M. & Woodward, G. Emerging horizons in biodiversity and ecosystem functioning research. *Trends Ecol. Evol.* **24,** 505–514 (2009).
15. Venail, P. A. *et al.* Diversity and productivity peak at intermediate dispersal rate in evolving metacommunities. *Nature* **452,** 210–214 (2008).
16. Levins, R. *Evolution in Changing Environments: Some Theoretical Considerations* (Princeton Univ. Press, 1968).
17. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19,** 207–233 (1988).
18. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15,** 173–190 (2002).
19. MacLean, R. C. Adaptive radiation in microbial microcosms. *J. Evol. Biol.* **18,** 1376–1386 (2005).
20. Buckling, A., Maclean, R. C., Brockhurst, M. A. & Colegrave, N. The Beagle in a bottle. *Nature* **457,** 824–829 (2009).
21. Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294,** 843–845 (2001).
22. Cardinale, B. J. *et al.* Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proc. Natl Acad. Sci. USA* **104,** 18123–18128 (2007).
23. Stachowicz, J. J., Graham, M., Bracken, M. E. S. & Szoboszlai, A. I. Diversity enhances cover and stability of seaweed assemblages: the role of heterogeneity and time. *Ecology* **89,** 3008–3019 (2008).
24. MacLean, R. C., Bell, G. & Rainey, P. B. The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens. Proc. Natl Acad. Sci. USA* **101,** 8072–8077 (2004).
25. Salles, J. F., Poly, F., Schmid, B. & Le Roux, X. Community niche predicts the functioning of denitrifying bacterial assemblages. *Ecology* **90,** 3324–3332 (2009).
26. Bell, T. *et al.* A linear model method for biodiversity-ecosystem functioning experiments. *Am. Nat.* **174,** 836–849 (2009).
27. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96,** 1463–1468 (1999).
28. Abrams, P. A. Adaptive change in the resource-exploitation traits of a generalist consumer: the evolution and coexistence of generalists and specialists. *Evolution* **60,** 427–439 (2006).
29. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9,** 1146–1156 (2006).
30. Bell, T. *et al.* in *Biodiversity, Ecosystem Functioning and Human Well-being* (eds Naeem, S. *et al.*) 121–133 (Oxford Univ. Press, 2009).

METHODS

Bacterial isolation. Bacterial strains were isolated from coastal sea water sampled from the Bay of Blanes, Spain (40° 40′ N, 2° 50′ E) on 20–21 September 2007. Five replicate samples of 100 µl of sea water were spread on marine agar plates (BD Difco Marine Agar 2216; autoclaved for 20 min at 121 °C) and grown for 5 d at 12 °C, which was the *in situ* temperature at sampling time. Ninety-five colonies with distinct morphotypes (that is, size, shape and colour) were isolated over four weeks, clean-streaked three times and frozen in glycerol at −80 °C.

We then sequenced the 16S rDNA genes of the ancestral strains to confirm different taxa. A single colony of the 95 isolates was picked and dissolved in 10 µl TE (10 nM Tris, 1mM EDTA, pH 8.0) buffer, heated for 5 min at 95 °C and centrifuged briefly. The supernatant (1 µl) was used as the PCR template for 16S rDNA gene amplification. The PCR reaction buffer (total volume, 25 µl) contained 200 µM of each deoxynucleoside triphosphate in 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM MgCl$_2$, 0.2 µM of primers 27F and 1492R[31] and ~2.5 units of puReTaq polymerase as included in the illustra puReTaq Ready-To-Go PCR beads kit (GE Healthcare). The PCR thermal cycling programme was as follows: 95 °C for 2 min; 30 cycles of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 45 s; and 72 °C for 7 min. The PCR products were sent to AGOWA Genomics for unidirectional sequencing using primers 27F and 519R[31]. The quality of the sequences was controlled by removing traces of the sequencing primers by using PHRED[32] with a base-calling score of $n \geq 20$. Ambiguous base calls (that is, 'N's) at the ends of the sequences were also trimmed away. All sequences were analysed using the programs MALLARD[33] and CHECK_CHIMERA from the Ribosomal Database Project[34]. Neither program detected any chimaeras. Resulting sequences were then compared with the SILVA database[35] using the program BLAST[36]. A phylogenetic tree was built to infer the relationships among the ancestral strains and their closest known relatives. In combination with their morphological characteristics, we considered the 95 selected strains as different taxa (Supplementary Information, section 6).

All 95 bacterial strains were tested for production of prophage (that is, to see whether they contained an inducible viral genome) on treatment with and without (control) the inducing agent mitomycin C at a final concentration of 1 µg ml^{-1} (ref 31). Incubations were carried out in 96-well microplates for 24 h. The growth kinetics of each strain was obtained by inoculating 10 µl of overnight cultures in 200 µl marine broth culture medium (BD Difco Marine Broth 2216; autoclaved for 20 min at 121 °C). Cultures were allowed to grow in the dark at 20 °C for 48 h. Changes in cell density were measured by the amount of light absorbance at 590-nm wavelength (FLUOstar OPTIMA spectrophotometer, BMG) every 15 min. We calculated the difference in the absorbance at stationary phase between control and treated samples, and found a bimodal distribution (no effect and strong effect of mitomycin C) with a threshold corresponding to a 30% growth reduction by mitomycin C. We defined a strain containing cells with an inducible viral genome as a strain with a growth reduction larger than this threshold, and therefore consider strains under this threshold to be phage free. Of those strains that fell below the 30% threshold we randomly selected 31 strains for the selection experiment.

Selection experiment. BIOLOG EcoPlates contain 31 different carbon substrates (plus one blank) belonging to different chemical families. In addition to the carbon substrates, each well contains a fluorogenic tetrazolium dye (5 cyano-2,3 ditolyl tetrazolium chloride), which is reduced to a violet-fluorescent formazan molecule when the carbon source is oxidized. Colour development was measured spectrophotometrically at 590 nm with a FLUOstar OPTIMA spectrophotometer and used as a proxy of metabolic activity[37]. Each of the 31 strains was used to establish three replicates of specialist and generalist selection lines. For the specialist treatment, each strain was assigned at random to one of the 31 carbon substrates. The EcoPlates used for the specialist treatment were prepared 2 h before each transfer with the addition of 140 µl of M9 minimal salts (0.1 g l^{-1} NH$_4$Cl, 6 g l^{-1} Na$_2$HPO$_4$, 3 g l^{-1} KH$_2$PO$_4$, 0.5 g l^{-1} NaCl) with salinity adjusted to 35.6 by the addition of NaCl, to match the salinity of the environment from which they were sampled. For the generalist treatment, we obtained a complex medium by mixing all of the 31 carbon EcoPlate substrates. The EcoPlates were prepared as for the specialists and after 30 min the contents of the EcoPlates (except the blanks) were transferred into a sterile flask, mixed with an orbital shaker and redistributed across a 96-well sterile microplate.

Each colony was initially grown for 24 h at 20 °C in 0.5 ml marine broth medium under constant orbital shaking. This solution (10 µl) was then used to inoculate each EcoPlate well from the two selection treatments. We intentionally did not wash the cells of marine broth medium to transfer a small quantity of this medium and assure the survival of the strains. Preliminary trials showed most of the strains did not initially survive in the EcoPlates without the marine broth. Consequently, for the first seven transfers we added a minimal quantity of marine broth, which we reduced from 7% (v/v) to 0% in steps of 1% at each transfer. The bacteria were incubated for 48 h in the dark at 20 °C in humid chambers. Bacteria were serially

transferred to maintain maximal growth rate and to renew the substrate. A transfer consisted of inoculating every well from a new plate with 10 µl of the corresponding previous well. Before each transfer, new EcoPlates were prepared as described above. The mixed medium was prepared at every tenth transfer and stocked at 4 °C. The selection experiment ran for 32 transfers (several hundreds generations). At the thirty-second transfer, the contents of every well was amended with glycerol (50% v/v) and frozen at −80 °C.

We measured light absorbance after conducting the selection experiment. Several lineages or replicates went extinct during the experiment, mostly for the specialist treatment. Therefore, we selected the most productive replicate among the lineages that survived the specialist treatment, and the corresponding lineages of the generalist treatment and the ancestors. We used 20 strains/lineages out of the 31 that were subject to the selection experiment for the BEF experiment.

BEF experiment. We assembled random combinations of species at six levels of species richness for each of the three treatments (ancestors, evolved specialists and generalists). We used an experimental design that allowed separation of the effects of species richness and species composition[9,26]. The experimental design consisted of a set of 20/s microcosms, each with s species present. Within this set, the microcosm assemblages were constructed by sampling all of the 20 species without replacement. The construction of a system of microcosms was carried out independently four times. We chose values of s to be every factor of 20 ($s = 1, 2, 4, 5, 10, 20$), so for any given s the number of assemblages considered was $4 \times 20/s$. Each assemblage was replicated three times, so in the experiment as a whole there were $3 \times 3 \times 4 \times (1 + 2 + 4 + 5 + 10 + 20) = 1,512$ microcosms.

Bacterial communities were assembled in six sterile 96-well, 1-ml microplates. Bacteria were initially grown for 24 h at 20 °C in 6 ml marine broth medium under constant orbital shaking in humid chambers. The cultures were centrifuged (5 min at 3,500 r.p.m.) and washed by eliminating the marine broth and adding 6 ml M9 minimal salts with buffered salinity. Because the cultures had different productivities in the marine broth, we first measured cell density by flow cytometry[38]. We adjusted the cell density to a concentration of 5×10^{-5} cells ml^{-1} with buffered M9. The bacteria were left in starvation for 2 h before 20 µl (40 µl for the monocultures) was inoculated into the appropriate wells. Once the cultures were distributed in the appropriate wells of the six plates, 10 µl of each community was transferred into three replicated microplates containing 140 µl of the mixed medium (that is, the assemblages were initiated with 5×10^{-3} cells), for a total of 18 microplates. The cultures were incubated at 20 °C in humid chambers for 72 h. Light absorbance at 590 nm was measured at 48 and 72 h.

Assays. Assays were conducted to measure strain/lineage performances on each carbon substrate at the end of the experiment. Before the assays, frozen cultures from the end of the selection period were reconditioned in 6 ml marine broth medium for 24 h at 20 °C in humid chambers under constant orbital shaking. The cultures were centrifuged (5 min at 3,500 r.p.m.) and washed by removing the marine broth and adding M9 minimal salts with buffered salinity to adjust cell density to 3.3×10^{-6} cells ml^{-1} (a pilot study showed that this concentration was optimal to obtain a signal differentiating strains). The cultures were left in starvation for 2 h. The EcoPlates were prepared with 120 µl of the buffered M9 solution. The EcoPlates were incubated with 30 µl of culture. Each strain/lineage was incubated in triplicate at 20 °C in humid chambers. Light absorbance at 590 nm was measured after 48 h.

Statistical analyses. The selection treatment (three levels) and species richness (log-transformed) were entered into an analysis of covariance of the bacterial productivity. The dependent variable was the light absorbance at 590 nm and we analysed the effects of the experimental treatments after 48 and 72 h. The results of the assays were averaged over the three replicates. A strain/lineage was considered to be able to exploit a carbon substrate when the light absorbance was larger than the 95% of the distribution of the blanks. For each assemblage, we calculated the NDI, which is the total number of carbon substrates that a community is able to exploit, calculated on the basis of the community composition and the individual ability of each strain/lineage to exploit the carbon substrates. A second analysis of covariance was conducted (excluding the ancestors) considering the NDI instead of species richness. We also fitted a linear model that assessed the effect of species richness and species identity on ecosystem functioning without requiring knowledge of the contribution of individual species to ecosystem functioning in mixture (see ref. 26 for details of this methodology). The model returns species-specific coefficients that could be interpreted as the contribution of individual species to ecosystem functioning relative to the average species.

31. Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 115–147 (Wiley, 1991).
32. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).

33. Ashelford, K. E. *et al.* New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* **72,** 5734–5741 (2006).
34. Maidak, B. L. *et al.* A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res.* **27,** 171–173 (1999).
35. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
36. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
37. Bochner, B. R. Sleuthing out bacterial identities. *Nature* **339,** 157–158 (1989).
38. Bouvier, T., del Giorgio, P. A. & Gasol, J. M. A comparative study of the cytometric characteristics of high and low nucleic-acid bacterioplankton cells from different aquatic ecosystems. *Environ. Microbiol.* **9,** 2050–2066 (2007).

LETTER

Rapid evolutionary innovation during an Archaean genetic expansion

Lawrence A. David[1] & Eric J. Alm[1,2,3]

The natural history of Precambrian life is still unknown because of the rarity of microbial fossils and biomarkers[1,2]. However, the composition of modern-day genomes may bear imprints of ancient biogeochemical events[3–6]. Here we use an explicit model of macro-evolution including gene birth, transfer, duplication and loss events to map the evolutionary history of 3,983 gene families across the three domains of life onto a geological timeline. Surprisingly, we find that a brief period of genetic innovation during the Archaean eon, which coincides with a rapid diversification of bacterial lineages, gave rise to 27% of major modern gene families. A functional analysis of genes born during this Archaean expansion reveals that they are likely to be involved in electron-transport and respiratory pathways. Genes arising after this expansion show increasing use of molecular oxygen ($P = 3.4 \times 10^{-8}$) and redox-sensitive transition metals and compounds, which is consistent with an increasingly oxygenating biosphere.

Describing the emergence of life on our planet is one of the grand challenges of the biological and Earth sciences. Yet the roughly three-billion-year history of life preceding the emergence of hard-shelled metazoans remains largely unknown[1]. So far, the best-understood event in early Earth history is the Great Oxidation Event, which is believed to have followed the development of oxygenic photosynthesis by ancestors of modern cyanobacteria[7] (although the precise timeline remains controversial[2,8]). If DNA sequences from extant organisms bear an imprint of this event, they can be used to make and test predictions; for example, genes that use molecular oxygen are more likely to appear in organisms that emerged after the Great Oxidation Event. However, the transfer of genes across species can obscure patterns of descent and disrupt our ability to correlate gene histories with the geochemical record[9]. For example, widely distributed genes may descend from a Last Universal Common Ancestor, as is widely believed to have occurred for the translational machinery[10], or they may have been dispersed by horizontal gene transfer (HGT)[11,12], as with antibiotic resistance cassettes.

We developed a new phylogenomic method, AnGST (analyser of gene and species trees), that explicitly accounts for HGT by comparing individual gene phylogenies with the phylogeny of organisms (the 'tree of life') and generated detailed evolutionary histories for 3,983 major gene families. Gene histories reveal marked changes in the rates of gene birth, gene duplication, gene loss and HGT over geological timescales (Fig. 1), including a burst of *de novo* gene-family birth between 3.33 and 2.85 Gyr ago, which we refer to as the Archaean Expansion. This window gave rise to 26.8% of extant gene families and coincides with a rapid bacterial cladogenesis (Supplementary Fig. 15). A spike in the rate of gene loss (about 3.1 Gyr ago) follows the expansion and may represent the consolidation of newly evolved phenotypes, as ancestral genomes became specialized for emerging niches. After 2.85 Gyr ago, the rates of both gene loss and gene transfer stabilized at roughly modern-day levels. The rates of *de novo* gene birth and duplication after the Archaean Expansion seem to show opposite trends: *de novo* gene-family birth rates decrease and duplication rates increase over

time. The near absence of *de novo* birth in modern times probably reflects the fact that ORFan gene families (gene families found in only a single genome), which are widespread across all major prokaryotic groups, are not considered in this study[13]. The excess of gene duplications and ORFans in modern genomes suggests that novel genes from both sources experience high turnover. Although we did not observe changes in the rate of HGT after the Archaean Expansion, we did detect an over-representation of HGT from α-proteobacteria to ancient eukaryotes ($P = 3.3 \times 10^{-7}$, Wilcoxon rank sum test) and from cyanobacteria to plants ($P = 8.3 \times 10^{-6}$, Wilcoxon rank sum test). These patterns of HGT probably reflect the endosymbioses that gave rise to mitochondria and chloroplasts[14,15], and serve to validate our phylogenomic approach.

What evolutionary factors were responsible for the period of innovation marked by the Archaean Expansion? Although we cannot provide an unequivocal answer to this question with the use of gene birth dates alone, we can ask whether the functions of genes born during this time suggest plausible hypotheses. In general, birth of metabolic genes was enriched during the expansion, and especially those involved in energy production and coenzyme metabolism (Supplementary Table 2); however, further inspection also reveals an enrichment for metabolic-gene-family birth before the Archaean Expansion. To focus on specific metabolic changes linked to the Archaean Expansion we first grouped genes according to the metabolites they used, and then directly compared the occurrence of these metabolites in genes born during the Archaean Expansion with their abundance before the Archaean Expansion. The results are striking: the metabolites specific to the Archaean Expansion (positive bars in Fig. 2 inset) include most of the compounds annotated as redox/e$^-$ transfer (blue bars), with Fe–S-binding, Fe-binding and O_2-binding gene families showing the most significant enrichment (false discovery rate < 5%, Fisher's exact test). Gene families that use ubiquinone and FAD (key metabolites in respiration pathways) are also enriched, albeit at slightly lower significance levels (false discovery rate < 10%). The ubiquitous NADH and NADPH are a notable exception to this trend and seem to have had a function early in life history. By contrast, enzymes linked to nucleotides (green bars) showed strong enrichment in genes of more ancient origin than the expansion.

The observed bias in metabolite use suggests that the Archaean Expansion was associated with an expansion in microbial respiratory and electron transport capabilities. Proving this association to be causal is beyond the power of our phylogenomic model. Yet this hypothesis is appealing because more efficient energy conservation pathways could increase the total free-energy budget available to the biosphere, possibly enabling the support of more complex ecosystems and a concomitant expansion of species and genetic diversity. We note, however, that although the use of oxygen as a terminal electron acceptor would have significantly increased biological energy budgets, oxygen-using genes were only enriched towards the end of the expansion (Supplementary Fig. 10). Thus, the earliest redox genes identified as part of the expansion were likely to have been used in anaerobic respiration or in oxygenic or
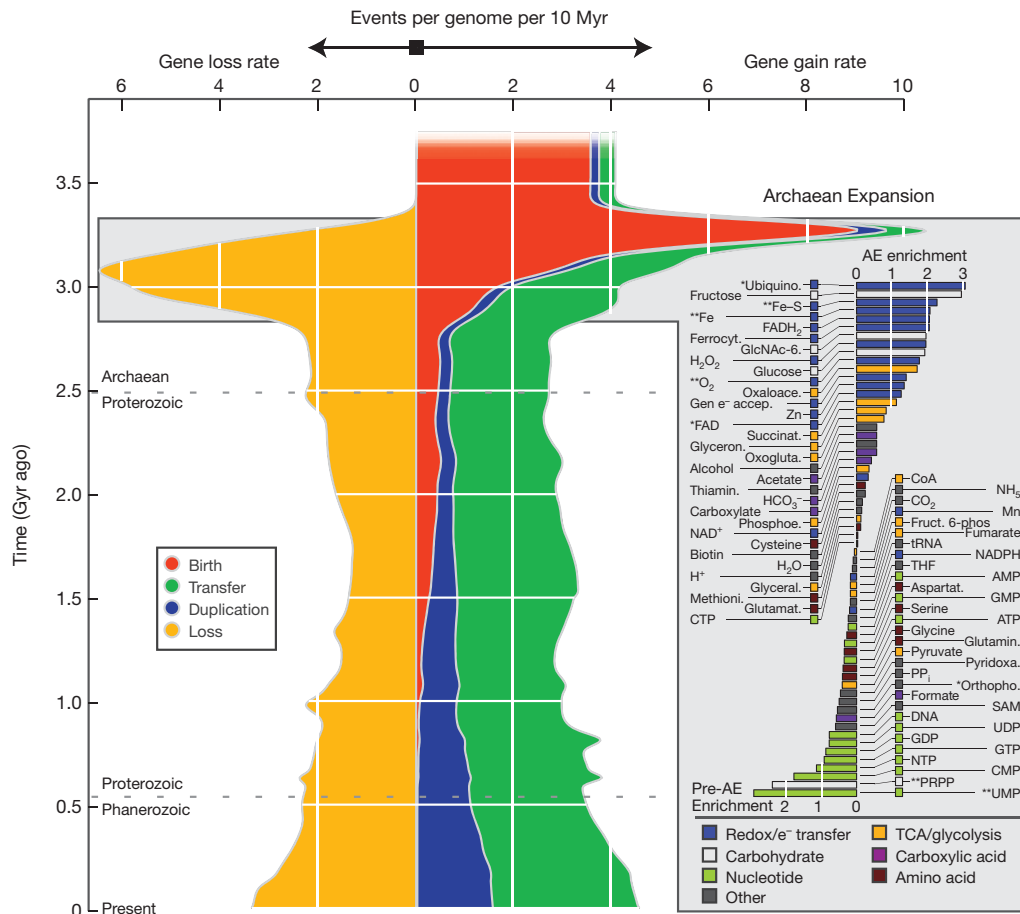
Figure 1 | Rates of macroevolutionary events over time. Average rates of gene birth (red), duplication (blue), HGT (green), and loss (yellow) per lineage (events per 10 Myr per lineage) are shown. Events that increase gene count are plotted to the right, and gene loss events are shown to the left. Genes already present at the Last Universal Common Ancestor are not included in the analysis of birth rates because the time over which those genes formed is not known. The Archaean Expansion (AE) was also detected when 30 alternative chronograms were considered (Supplementary Fig. 9). The inset shows metabolites or classes of metabolites ordered according to the number of gene families that use them that were born during the Archaean Expansion compared with the number born before the expansion, plotted on a log₂ scale. Metabolites whose enrichments are statistically significant at a false discovery rate of less than 10% or less than 5% (Fisher's Exact Test) are identified with one or two asterisks, respectively. Bars are coloured by functional annotation or compound type (functional annotations were assigned manually). Metabolites were obtained from the KEGG database release 51.0 (ref. 27) and associated with clusters of orthologous groups of proteins (COGs) using the MicrobesOnline September 2008 database[28]. Metabolites associated with fewer than 20 COGs or sharing more than two-thirds of gene families with other included metabolites are omitted. Abbreviations are defined in Supplementary Table 3.

anoxygenic photosynthesis and may have been co-opted later for use in aerobic respiration pathways.

Our metabolic analysis supports an increasingly oxygenated biosphere after the Archaean Expansion, because the fraction of proteins using oxygen gradually increased from the expansion to the present day (Fig. 2; $P = 3.4 \times 10^{-8}$, two-sided Kolmogorov–Smirnov test). Further indirect evidence of increasing oxygen levels comes from compounds whose availability is sensitive to global redox potential. We observe significant increases over time in the use of the transition metals copper and molybdenum (Fig. 2; false discovery rate < 5%, two-sided Kolmogorov–Smirnov test), which is in agreement with geochemical models of these metals' solubility in increasingly oxidizing oceans[5,6] and with molybdenum enrichments from black shales suggesting that molybdenum began accumulating in the oceans only after the Archaean eon[16]. Our prediction of a significant increase in nickel utilization accords with geochemical models that predict a tenfold increase in the concentration of dissolved nickel between the Proterozoic eon and the present day[5] but conflicts with a recent analysis of banded iron formations that inferred monotonically decreasing maximum concentrations of dissolved nickel from the Archaean onwards[17]. The abundance of enzymes using oxidized forms of nitrogen (N_2O and NO_3) also grows significantly over time, with one-third of nitrate-binding gene

families appearing at the beginning of the expansion and three-quarters of nitrous-oxide-binding gene families appearing by the end of the expansion. The timing of these gene-family births provides phylogenomic evidence for an aerobic nitrogen cycle by the Late Archaean[18].

However, one striking discrepancy between our phylogenomic patterns and geochemical predictions is a modest but significant increase in iron-using genes over time (Fig. 2; false discovery rate < 5%, two-sided Kolmogorov–Smirnov test). Declining iron solubility in oxygenated ocean surface waters and sulphide-mediated removal of iron from anoxic deeper waters are thought to have decreased overall iron bioavailability during the Proterozoic[19]. If the abundance of iron-using genes tracks iron bioavailability, we would expect these genes to decrease in abundance after the Archaean. The conflicting phylogenomic result may reflect the confounding effect of evolutionary inertia, whereby microbes could have found more success in evolving a handful of metal-acquisition proteins (for example siderophores) rather than replacing a host of iron-binding proteins in the face of declining iron availability[5]. Alternatively, the insolubility of iron in modern oceans may be offset by large organic pools of iron.

Our chronologies of oxygen and redox-sensitive metal and compound utilization suggest ancient increases in oxygen bioavailability, as well as an Archaean biosphere with some of the basic genetic components
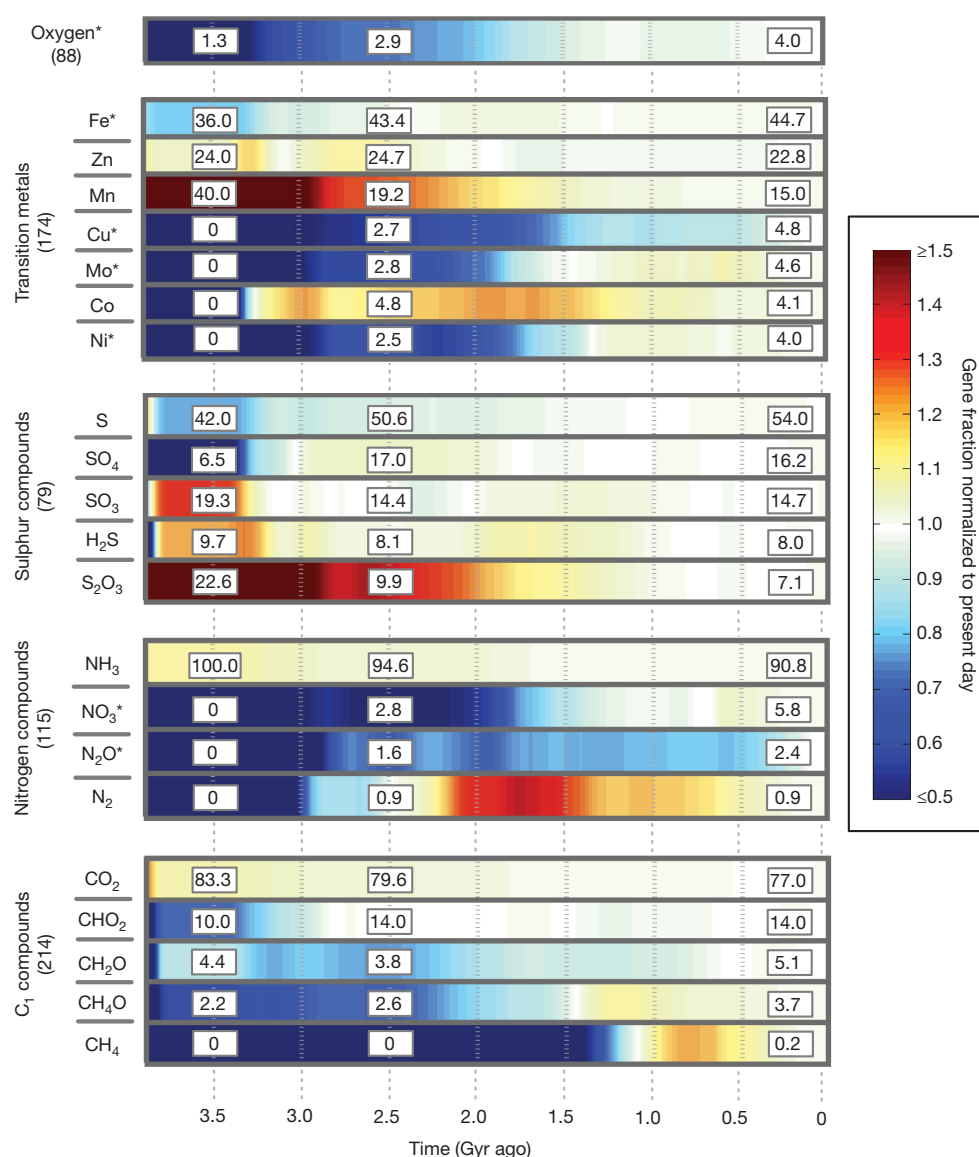
Figure 2 | Genome utilization of redox-sensitive compounds over time. The top panel illustrates a gradual increase in the fraction of enzymes that bind molecular oxygen predicted to be present over Earth history ($P = 3.4 \times 10^{-8}$, two-sided Kolmogorov–Smirnov test). Colours indicate abundance normalized to present-day values. The lower four panels group transition metals, nitrogen compounds, sulphur compounds and C_1 compounds. The fraction of each group's associated genes that bind a given compound, normalized to present-day fractions, is shown over time with a colour gradient. Enclosed boxes show raw fractional values at three time points: 3.5 Gyr ago (left), 2.5 Gyr ago (middle) and the present day (right). For example, 19.2% of transition-metal-binding genes are predicted to have bound Mn 2.5 Gyr ago, a value 1.28-fold the size of the present-day percentage of 15.0%. Values within parentheses give the overall number of gene families in each group. To determine which compounds showed divergent genome utilization over time, the timing of copy number changes for each compound's associated genes was compared with a background model derived from all other compounds. Compounds whose utilization significantly differs from the background model are marked with an asterisk (false discovery rate < 5%, two-sided Kolmogorov–Smirnov test). Nitrite and nitric oxide are not shown, because of their COG-binding similarity to nitrate and nitrous oxide, respectively.

required for oxygenic photosynthesis and respiration. These results are consistent with recent biomarker-based evidence for oxygenesis preceding the Palaeoproterozoic era by hundreds of millions of years[20]. Still, a precise timeline for the origins of oxygenesis is currently beyond the resolution of our phylogenomic model. In the results described above, we estimated lineage divergence times with PhyloBayes[21], which enabled us to explicitly account for uncertainty in the timing of inferred events (Supplementary Fig. 13). An alternative model of evolutionary rates[22] dated the rapid bacterial cladogenesis to 2.75–2.5 Gyr ago (in contrast to 3.33–2.85 Gyr for PhyloBayes) but still finds evidence for an Archaean Expansion (Supplementary Fig. 9) characterized by the emergence of electron transport genes. Uncertainty or errors in the reference tree may further decrease the power of our phylogenomic model, obscuring evidence for all except the most extreme geochemical events. Future studies that benchmark biomarker and other geochemical data against

the predicted age of associated gene families could be used to test and refine the 'tree of life', ultimately yielding an abundant and reliable source of Precambrian fossils: modern-day genomes.

METHODS SUMMARY

We developed AnGST to account for gene transfer, duplication, loss and *de novo* birth by comparing individual gene phylogenies with a previously described reference phylogeny[23]. We refer to this process as tree reconciliation and provide a detailed description of the AnGST algorithm in Supplementary Methods. Unlike some previous methods[24–26], AnGST uses the topology of the gene family tree rather than just its presence or absence across genomes and can infer the direction of gene transfer in addition to gene duplication, birth and loss events. AnGST also accounts for uncertainty in gene trees by incorporating reconciliation into the tree-building process: the tree that minimizes the macroevolutionary cost function but is still supported by the sequence data is chosen as the best gene tree. To assess the sensitivity of our method to the reference tree topology, we reconciled gene families

against 30 alternative reference trees rooted on either the bacterial, archaeal or eukaryotic branches. Inferred gene-family birth ages were consistent across the ensemble of reference trees, and the Archaean Expansion was a uniformly observed feature (Supplementary Figs 8 and 9). A conservative set of eight temporal constraints was selected from the geochemical and palaeontological literature (Supplementary Fig. 7 and Supplementary Table 1), and the PhyloBayes software package was used to infer a range of divergence times for each ancestral lineage on the reference tree[21]. We did not apply temporal constraints to lineage ages on the gene trees.

1. Nisbet, E. G. & Sleep, N. H. The habitat and nature of early life. *Nature* **409,** 1083–1091 (2001).
2. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455,** 1101–1104 (2008).
3. Dupont, C. L., Yang, S., Palenik, B. & Bourne, P. E. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proc. Natl Acad. Sci. USA* **103,** 17822–17827 (2006).
4. Dupont, C. L., Butcher, A., Valas, R. E., Bourne, P. E. & Caetano-Anollés, G. History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc. Natl Acad. Sci. USA* **107,** 10567–10572 (2010).
5. Saito, M. A., Sigman, D. M. & Morel, F. M. M. The bioinorganic chemistry of the ancient ocean: the co-evolution of cyanobacterial metal requirements and biogeochemical cycles at the Archean–Proterozoic boundary? *Inorg. Chim. Acta* **356,** 308–318 (2003).
6. Zerkle, A. L., House, C. H. & Brantley, S. L. Biogeochemical signatures through time as inferred from whole microbial genomes. *Am. J. Sci.* **305,** 467–502 (2005).
7. De Marais, D. J. When did photosynthesis emerge on Earth? *Science* **289,** 1703–1705 (2000).
8. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archean molecular fossils and the early rise of eukaryotes. *Science* **285,** 1033–1036 (1999).
9. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19,** 2226–2238 (2002).
10. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96,** 3801–3806 (1999).
11. Ragan, M. A. & Beiko, R. G. Lateral genetic transfer: open issues. *Phil. Trans. R. Soc. Lond. B* **364,** 2241–2251 (2009).
12. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).
13. Fischer, D. & Eisenberg, D. Finding families for genomic ORFans. *Bioinformatics* **15,** 759–762 (1999).
14. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82,** 4443–4447 (1985).
15. Giovannoni, S. J. *et al.* Evolutionary relationships among cyanobacteria and green chloroplasts. *J. Bacteriol.* **170,** 3584–3592 (1988).
16. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452,** 456–459 (2008).
17. Konhauser, K. O. *et al.* Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* **458,** 750–753 (2009).
18. Garvin, J., Buick, R., Anbar, A. D., Arnold, G. L. & Kaufman, A. J. Isotopic evidence for an aerobic nitrogen cycle in the latest Archean. *Science* **323,** 1045–1048 (2009).
19. Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396,** 450–453 (1998).
20. Waldbauer, J. R., Sherman, L. S., Sumner, D. Y. & Summons, R. E. Late Archean molecular fossils from the Transvaal Supergroup record the antiquity of microbial diversity and aerobiosis. *Precambr. Res.* **169,** 28–47 (2009).
21. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21,** 1095–1109 (2004).
22. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19,** 301–302 (2003).
23. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311,** 1283–1287 (2006).
24. Alm, E., Huang, K. & Arkin, A. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLOS Comput. Biol.* **2,** e143 (2006).
25. Kunin, V. & Ouzounis, C. A. The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* **13,** 1589–1594 (2003).
26. Snel, B., Bork, P. & Huynen, M. A. Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* **12,** 17–25 (2002).
27. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
28. Alm, E. J. *et al.* The MicrobesOnline Web site for comparative genomics. *Genome Res.* **15,** 1015–1022 (2005).

LETTER

Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs

Calvin H. Jan[1,2], Robin C. Friedman[1,2,3], J. Graham Ruby[1,2]† & David P. Bartel[1,2]

Post-transcriptional gene regulation frequently occurs through elements in mRNA 3′ untranslated regions (UTRs)[1,2]. Although crucial roles for 3′UTR-mediated gene regulation have been found in *Caenorhabditis elegans*[3–5], most *C. elegans* genes have lacked annotated 3′UTRs[6,7]. Here we describe a high-throughput method for reliable identification of polyadenylated RNA termini, and we apply this method, called poly(A)-position profiling by sequencing (3P-Seq), to determine *C. elegans* 3′UTRs. Compared to standard methods also recently applied to *C. elegans* UTRs[8], 3P-Seq identified 8,580 additional UTRs while excluding thousands of shorter UTR isoforms that do not seem to be authentic. Analysis of this expanded and corrected data set suggested that the high A/U content of *C. elegans* 3′UTRs facilitated genome compaction, because the elements specifying cleavage and polyadenylation, which are A/U rich, can more readily emerge in A/U-rich regions. Indeed, 30% of the protein-coding genes have mRNAs with alternative, partially overlapping end regions that generate another 10,480 cleavage and polyadenylation sites that had gone largely unnoticed and represent potential evolutionary intermediates of progressive UTR shortening. Moreover, a third of the convergently transcribed genes use palindromic arrangements of bidirectional elements to specify UTRs with convergent overlap, which also contributes to genome compaction by eliminating regions between genes. Although nematode 3′UTRs have median length only one-sixth that of mammalian 3′UTRs, they have twice the density of conserved microRNA sites, in part because additional types of seed-complementary sites are preferentially conserved. These findings reveal the influence of cleavage and polyadenylation on the evolution of genome architecture and provide resources for studying post-transcriptional gene regulation.

We developed a high-throughput method to identify 3′ ends of mRNAs and other polyadenylated transcripts (Fig. 1a). This method, called poly(A)-position profiling by sequencing (3P-Seq), begins with a splint-ligation that favours ends of poly(A) tails when appending a biotinylated primer-binding site (Fig. 1a, step 1). After partial digestion with T1 nuclease (which cuts after Gs; step 2), the polyadenylated ends are captured (step 3), and the poly(A) tail is reverse transcribed with dTTP as the only deoxynucleoside triphosphate (step 4). Digestion with RNase H releases the polyadenylated ends (step 5), which are purified (step 6) and prepared for high-throughput sequencing (step 7).

3P-Seq was designed to identify the 3′ ends of polyadenylated RNAs without recourse to oligo(dT) priming. Oligo(dT) priming can prime on internal A-rich regions of transcripts, thereby yielding artefacts difficult to distinguish from authentic polyadenylated transcripts because the artefacts also have untemplated As[9]. Although untemplated adenylates at the ends of 3P tags could not have arisen from internal-priming artefacts, in principle, such nucleotides could have arisen from polymerase/sequencing errors. Countering this possibility was the observation that homopolymeric runs containing untemplated

nucleotides at the ends of candidate 3P tags were overwhelmingly As (Fig. 1b). Thus, non-genomic terminal adenylates at the ends of 3P tags (a beneficial consequence of incomplete RNase H digestion near duplex termini (Fig. 1a)) provided compelling evidence that they derived from distal ends of bona fide polyadenylated transcripts.

To ensure proper assignment to polyadenylated transcripts, we considered as 3P tags only reads that both mapped uniquely to the genome and possessed at least two 3′-terminal adenylates, of which at least one was untemplated. Nearly 32 million reads from *C. elegans* met these criteria, including millions from each major developmental stage (embryo, L1, L2, L3, L4, adult) as well as dauer L3 worms and germ-line-deficient *glp-4(bn2)* mutant adults (Supplementary Table 1).
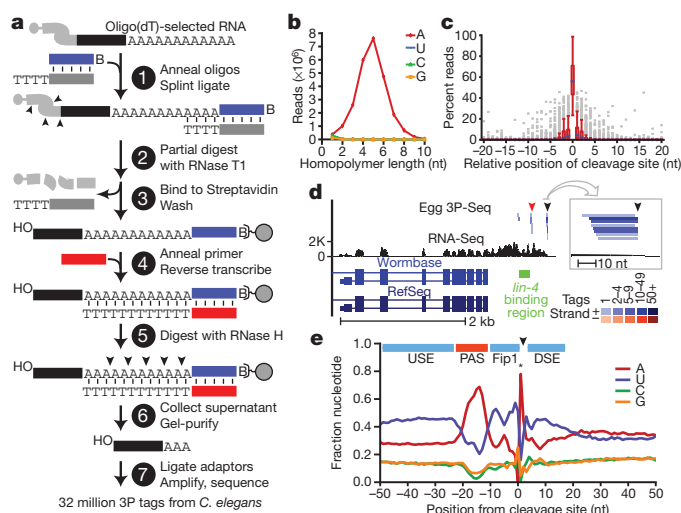

Figure 1 | Identification of *C. elegans* 3′UTRs. a, Schematic of the 3P-Seq protocol. See text for description. **b**, Sequence composition of homopolymer runs that were found at 3′ termini of candidate 3P tags and included ≥1 untemplated nucleotide. **c**, Cleavage heterogeneity surrounding the most abundant cleavage site (position 0). Box plots show results for 380 cleavage sites that were both between two non-A residues (which enabled precise mapping) and within the top quintile of 3P-tag abundance. **d**, The *lin-14* 3′UTRs. 3P tags from egg were mapped relative to RNA-Seq data[10], prior mRNA annotations from the indicated databases[6,11], and the proposed *lin-4*-binding region[4]. Distal and proximal cleavage sites are indicated (black and red arrowheads, respectively). A 50-nucleotide region containing the distal 3P cluster is enlarged (box). Each tag sequence with a unique genome match is depicted as a bar, coloured by tag frequency (key). **e**, Nucleotide sequence composition at mRNA end regions. Shown above are elements implicated in cleavage and polyadenylation (Supplementary Fig. 3c)[30], with colours reflecting their nucleotide composition (A-rich, red; U-rich, blue). The sharp adenosine peak at position +1 (*) was due only partly to cleavage before an A. Also contributing to this peak (and to both depletion of A at position −1 and blurring of sequence composition at other positions) was cleavage after an A, for which the templated A was assigned to the poly(A) tail, resulting in a −1 nucleotide offset from the cleavage-site register.

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. [2]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present address: Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California 94158, USA (J.G.R.).

Microheterogeneity at the cleavage and polyadenylation sites (hereafter called cleavage sites) often produced clusters of related 3P tags (Fig. 1c, d). All tags ending within 10 nucleotides of the most frequently implicated cleavage site were consolidated into a cluster, with this candidate cleavage site representing that of the cluster (Supplementary Data 1). Candidate sites were classified as mRNA cleavage sites if they were bridged by RNA-Seq reads[10] to the stop codons of RefSeq mRNAs[6,11], as illustrated for *lin-14* (Fig. 1d).

3P-Seq identified 24,033 distinct 3′UTRs, including at least one UTR for 16,261 (83%) of the RefSeq mRNAs (Supplementary Data 2). For 5,331 mRNAs, we revised the longest-isoform annotation by >10 nucleotides (usually by extending it), and for 5,852 mRNAs without 3′UTR annotations, we identified a UTR (Supplementary Fig. 1a). A parallel effort within the modENCODE project used oligo(dT)-based methods to also generate a greatly expanded data set of *C. elegans* 3′UTRs[8]; 8,580 of the 24,033 UTRs identified by 3P-Seq were not identified in that study (Supplementary Table 2). Our data were shared with the modENCODE consortium, thereby enabling them to annotate 8,758 novel UTRs that the oligo(dT)-based methods missed ((L. Hillier and R. Waterston, personal communication), Supplementary Table 3). Of the 3,280 RefSeq mRNAs not assigned 3′UTRs using 3P-Seq, most were from predicted genes without evidence of expression (Supplementary Fig. 1b). Of the remainder, most were expressed at extremely low levels (Supplementary Fig. 1c). We estimated that only 124 ± 56 (95% confidence interval) sites were missed by requiring that tags have an untemplated A (Supplementary Fig. 1d). Most histone mRNAs were assigned 3′UTRs, consistent with oligo(dT)-based results[8], but the polyadenylated forms of these mRNAs did not accumulate to levels detectable on RNA blots (Supplementary Fig. 2).

Apart from the A-rich segment corresponding to the polyadenylation signal (PAS) AAUAAA and its close variants (Supplementary Fig. 3a, b and Supplementary Table 4), the mRNA end regions were U-rich, presumably a feature of the binding sites of factors that enhance cleavage and polyadenylation (Fig. 1e and Supplementary Fig. 3c). Indeed, end regions that lacked a common PAS had exaggerated U-rich features surrounding an A-rich segment located where the PAS normally occurs (Supplementary Fig. 3d), which suggests that appropriate U-rich context can compensate for lack of a strong PAS[12].

3P-Seq was particularly useful for reliably identifying alternative UTR isoforms. Genes with tandem 3′UTRs possess proximal cleavage sites that, when used, create a shorter UTR that is a subfragment of longer versions (Fig. 2a). When identifying these shorter isoforms, we required that (1) the proximal site be represented by ≥2 independent 3P tags, (2) that these tags constitute ≥1% of the tags mapping between the distal site and the stop codon, and (3) the site be in an end region non-overlapping with that of a more distal site (that is, that the two cleavage sites be ≥40 nucleotides apart). These criteria identified 7,795 shorter isoforms, which corresponded to 31% of the Entrez genes with 3P-supported UTRs (Fig. 2a). As expected for sites sometimes

bypassed by the cleavage and polyadenylation machinery to allow production of longer isoforms, a larger fraction lacked a common PAS (Fig. 2b). Although less conserved than PASs for distal-most sites, PASs for proximal sites were more conserved than expected by chance (Supplementary Fig. 4b). Proximal isoforms had lengths typical of *C. elegans* UTRs, whereas distal isoforms were longer than typical UTRs (Supplementary Fig. 4a; $P < 10^{-300}$, Wilcoxon rank-sum test), hinting at even more elaborate UTR-mediated regulation.

Oligo(dT)-based results have been interpreted to show that a large class of proximal isoforms lack PASs and instead have A-rich regions immediately following their cleavage sites[8]. 3P-Seq, which avoids oligo(dT) priming, provided no evidence for this novel class of isoforms, suggesting that it is composed of false-positives that arose from internal priming on A-rich UTR regions, as illustrated for the *ubc-18* 3′UTR and confirmed by an RNase-protection experiment (Supplementary Fig. 5a, b). Of the 5,728 proposed cleavage sites supported by oligo(dT)-based methods but not 3P-Seq, 3,900 were sites of putative proximal isoforms (Supplementary Table 3), of which ~70% seem to have resulted from internal-priming artefacts (Supplementary Fig. 5c).

Genes with alternative last exons (ALEs) generate messages with completely different UTRs (Fig. 2a). We identified 1,398 ALEs distributed across 1,277 Entrez genes. Previous methods identified <25% of these ALEs (Supplementary Fig. 5d), presumably because data acquisition or analyses had focused on regions downstream of annotated stop codons[8], which illustrates advantages of 3P-Seq for identifying unanticipated UTRs. The PAS motifs and nucleotide composition associated with proximal ALE ends were comparable to those at distal ends (Fig. 2a, b), and the distal isoforms tended to be longer than both proximal isoforms and single UTRs (Supplementary Fig. 4c, $P < 10^{-5}$ and $<10^{-14}$, respectively, Wilcoxon rank-sum test).

Our analyses also identified a novel gene architecture, called the 'alternative operon'. *C. elegans* operons are each arrays of genes transcribed from a single promoter and split into separate mRNAs through the biochemically coupled processes of 3′-end formation and trans-splicing to splicing leader 2 (SL2)[13]. Reasoning that this coupling could result in SL2 trans-splicing to 3′-splice sites downstream of ALEs, we searched for a gene structure that differed from the canonical operon by a splice junction bridging exons from different genes of an operon (Fig. 2c). This search identified 12 alternative operons, including the *smg-6* locus (Supplementary Fig. 6 and Supplementary Table 5).

Among representative metazoans, *C. elegans* had the shortest 3′UTRs, with a length distribution approaching that of *Saccharomyces cerevisiae* (Fig. 3a) and a median length of 130 nucleotides, only one sixth that of human. *C. elegans* 3′UTRs were also the most A/U rich. Shorter UTRs tended to be the most A/U rich (Fig. 3b), and even after masking the UTR end regions, which are exceptionally U/A rich, a cross-species comparison revealed a significant inverse correlation between 3′UTR length and 3′UTR A/U content ($P = 0.0003$, $r^2 = 0.92$, Pearson
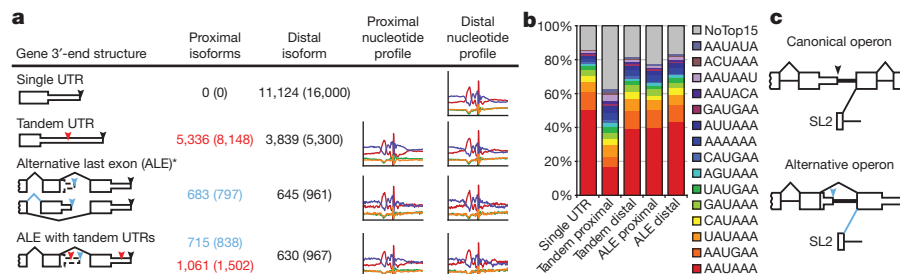

Figure 2 | Alternative 3′UTRs in *C. elegans*. a, Distribution of the 24,033 3P-Seq–supported UTRs among the types of alternative isoforms. For genes with ALEs that have tandem isoforms (bottom), the ALE tally indicates the number of distal isoforms of proximal ALEs (blue) and the tandem tally indicates the proximal tandem isoforms of all ALEs (red). In all cases, the distal isoform is the 3′-most cleavage site for each gene (black arrowhead). Also depicted are proximal tandem sites and proximal ALE sites (red and blue arrowheads,

respectively). Listed (in parenthesis) is the number of cleavage sites associated with each isoform type for the 34,513 3P-Seq-supported cleavage sites (which exceeded the number of unique UTRs because OERs produced multiple cleavage sites for the same UTR). The nucleotide composition near proximal and distal sites is shown (right). **b**, Frequency of PAS motifs for isoform types indicated. **c**, Schematics of canonical and alternative operons.
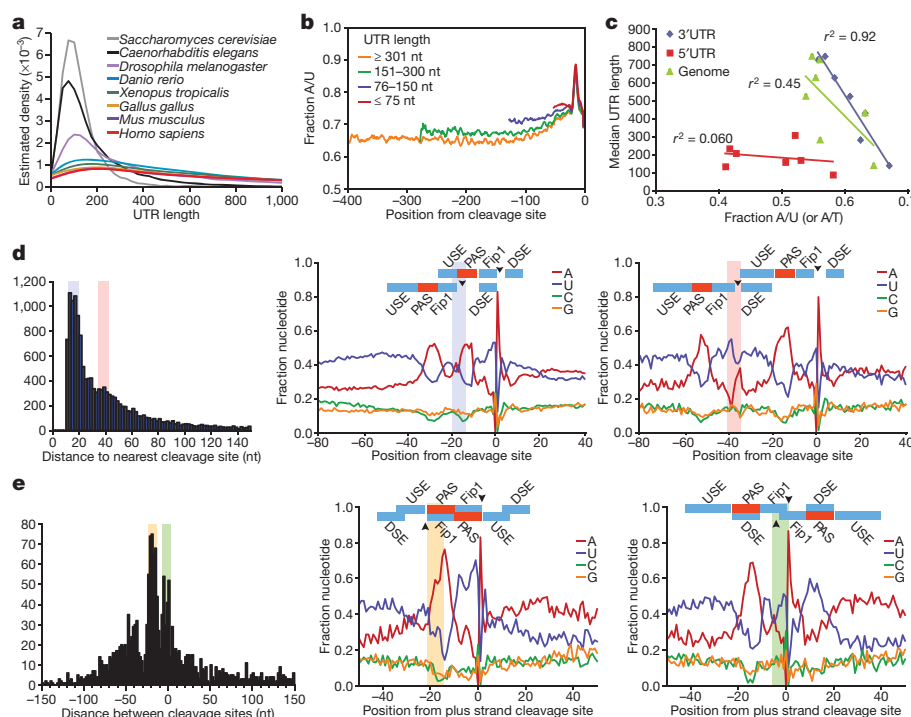
Figure 3 | Evolution and topology of 3′-end formation. a, 3′UTR length distributions for the indicated species, considering the most distal annotated isoform for each gene. **b**, A/U content for *C. elegans* 3′UTRs of the indicated lengths. **c**, Relationship between 3′UTR length and 3′UTR A/U content (disregarding content of the last 40 UTR nucleotides), 3′UTR length and genomic A/T content, and 5′UTR length and 5′UTR A/U content for the metazoan species in (**a**) (r^2, Pearson correlation coefficients). **d**, OERs. Distances between neighbouring cleavage sites are plotted (left). For peaks in the distribution at 15–20 and 35–40 nucleotides (shaded), nucleotide compositions of OERs are shown (middle and right, respectively), with proposed RNA-recognition elements coloured as in Fig. 1e. Arrowheads indicate cleavage sites, with shading also indicating positions of upstream cleavage. **e**, Convergent UTR overlap. Distances between convergent 3′ ends are plotted (left), with negative values indicating overlap. For peaks at 15–22 and (−2)–8 nucleotides of overlap (shaded), nucleotide compositions are shown (middle and right, respectively) as in (**d**), with shading indicating positions of minus-strand cleavage.

correlation), whereas correlations between either 5′UTR length and A/U content or 3′UTR length and genomic A/T content were less significant ($P = 0.30$ and 0.05, respectively; Fig. 3c). We speculate that this strong inverse correlation is causal; that is, higher A/U content favours the emergence of A/U-rich motifs that create proximal mRNA ends within existing 3′UTRs, thereby generating progressively shorter UTRs.

Also potentially related to progressive UTR shortening were the 7,116 UTRs with ≥2 closely spaced alternative cleavage sites. We did not classify these as tandem UTRs because the cleavage sites were very close to each other (<40 nucleotides, usually 12–22 nucleotides), implying overlapping end regions (OERs). This overlap tended to be phased, such that U-rich *cis*-acting elements could serve dual functions, binding alternative factors, depending on which cleavage site was being recognized (Fig. 3d). Although previous studies do not distinguish these isoforms from the heterogeneity normally found at UTR ends, proximal and distal OER isoforms were distinct in that each tended to have their own A-rich PASs (Fig. 3d). The 10,480 additional cleavage sites from OERs thus represented the largest class of alternative mRNA isoforms in *C. elegans* (Fig. 2a, compare UTR tallies with cleavage-site tallies).

The few additional nucleotides of distal OER isoforms presumably are dedicated to end recognition and processing (Fig. 3d), leaving little space for regulatory sites that could impart differential regulation. Thus, the importance of the OER isoforms might pertain instead to UTR evolution. The potential of the U-rich regions to serve dual functions would favour the emergence of new cleavage sites with OERs. Moreover, the higher A/U content of *C. elegans* UTRs compared to that of intergenic regions would favour the emergence of more upstream sites than downstream sites, which in turn could lead to progressive UTR shortening as the original signals acquire mutations

rendering them less able to compete for factors (Supplementary Fig. 7a). If nematode UTRs had a propensity to drift towards a minimum UTR length, longer UTRs, which have avoided this shortening, might display more evidence of cleavage-site retention. Indeed, the PASs of long UTRs were more frequently conserved than were those of shorter UTRs (Supplementary Fig. 7b, $P < 10^{-15}$, Kolmogorov–Smirnov test).

The alternating U- and A-rich elements defining UTR end regions provided opportunity for motifs to also serve double duty on opposite strands. Indeed, overlap of convergent UTRs occurred with a tri-modal distribution peaking at 5, 20 and 40 nucleotides, in which the A-rich PASs of each strand reciprocally served as U-rich motifs of the other strand (Fig. 3e). The bidirectionality of these composite sites was often selectively maintained (Supplementary Fig. 8a). Previously, a single peak in the distribution was observed at ~20 nucleotides of overlap, which was attributed to selective pressure to avoid RNAi[8]. Our data indicated a more complex overlap distribution that is better explained by preferential emergence of end regions where end elements of a convergent gene already provide some of the alternating A- and U-rich segments needed for end recognition. Although more extensive overlap can act to enforce mutually exclusive transcriptional regulation[14], expression of the overlapping gene pairs were no less correlated than were random pairs (Supplementary Fig. 8b). Hence, gene topology using palindromic arrangement of bidirectional elements provides a mechanism for genome compaction, effectively minimizing intergenic space downstream of 2,448 genes (a sixth of all genes with 3P-Seq-identified ends) without significantly impacting their regulatory autonomy.

Before considering targeting of the newly annotated 3′UTRs by microRNAs (miRNAs), we updated the set of confidently identified miRNAs using ~23 million genome-matching small-RNA sequences[15].
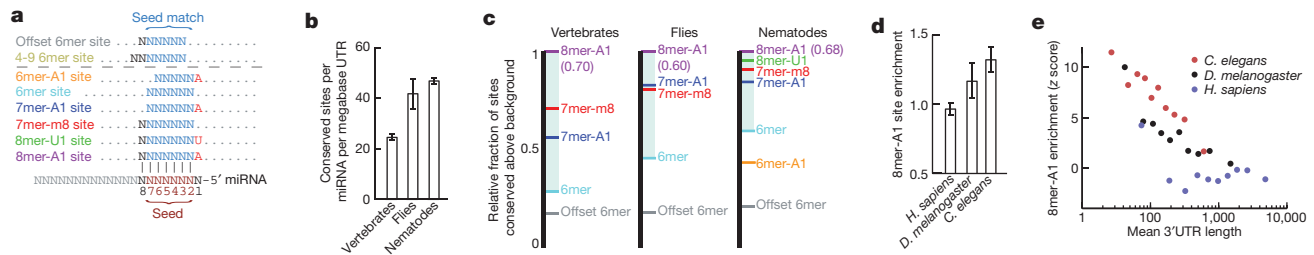
Figure 4 | MicroRNA targeting. a, Expanded repertoire of seed-matched sites preferentially conserved in nematode 3′UTRs. Sites conserved only marginally above chance are above the dashed line. Watson–Crick-matched residues, blue or black; residues independent of the miRNA sequence, red. **b**, Density of miRNA sites conserved above background, combining all site types at the maximally sensitive cutoff. Error bars, one standard deviation (calculated by repeating the analysis for each site type 50 times, each time using a different cohort of control sequences that matched the properties of the miRNA sequences[18]). **c**, Relative strength of miRNA site types across clades. Within each clade, two species of comparable divergence were selected. For each miRNA site type, the fraction of sites conserved above background in the two species was normalized to that of the 8mer-A1 (shown in parentheses). **d**, Enrichment of 8mer-A1 3′UTR sites above expectation based on dinucleotide content. Error bars, one standard deviation, derived as in (**b**). **e**, Relationship between 3′UTR length and site enrichment. Site enrichment is plotted for 3′UTRs of the indicated species sorted by length into ten equally sized bins.

Methods shown to identify miRNAs reliably in mammals[16] provided confident support for 147 annotated genes and 12 additional genes (Supplementary Table 6, Supplementary Fig. 9 and Supplementary Text). Five of the newly identified miRNAs derived from mirtrons (Supplementary Table 6), which are spliced and debranched introns that fold into pre-miRNA hairpins, thereby bypassing Drosha processing[17]. Although mirtrons are typically thought to be spliced from pre-mRNAs, three newly identified mirtrons and two pre-miRNAs reclassified as mirtrons (*mir-255* and *mir-2220*) are derived from host transcripts that did not seem to be protein coding (Supplementary Fig. 10 and Supplementary Table 6). We also generated developmental expression profiles for the 159 confidently annotated genes (Supplementary Fig. 11, Supplementary Tables 6 and 7 and Supplementary Text).

Methods used previously to detect miRNA site conservation in vertebrate genome alignments[18] found six types of preferentially conserved sites that matched the miRNA seed region (Fig. 4a), including an octamer site (8mer-U1) and a hexamer site (6mer-A1) not observed in vertebrates (Supplementary Figs 12a–c, 13 and Supplementary text). Efficacy of these six types was confirmed using two large-scale experimental data sets: mRNA fragments crosslinked to *C. elegans* miRNA silencing complexes[19] and mRNA changes in *C. elegans* miR-124 mutants[20] (Supplementary Fig. 12d and Supplementary Table 8).

Summing results for all six site types indicted that *C. elegans* UTRs have at least 9,093 ± 146 (95% confidence interval) selectively maintained miRNA sites, and that at least 27.4% ± 4.8% of the *C. elegans* 3′UTRs have been under selective pressure to retain miRNA targeting (Supplementary Fig. 12b, c). This percentage was nearly threefold greater than that detected previously in nematodes[21] and about half that observed for human UTRs[18], despite substantially shorter lengths of nematode UTRs, fewer nematode genomes available, and fewer conserved miRNA families in nematodes (60, compared to 87 in vertebrates). As in vertebrates[18], few preferentially conserved sites had mismatches or wobbles to the seed nucleotides (Supplementary Fig. 14). Indeed, the three most compelling sites with seed mismatches (two *let-7* sites in *lin-41*; and one *let-7* site in *hbl-1*; Supplementary Table 9) had all been implicated by earlier genetic studies[22,23]. The updated miRNA target predictions will be presented in TargetScanWorm, release 5.2 (targetscan.org).

Compared to human 3′UTRs, *C. elegans* 3′UTRs had twice the density of selectively conserved miRNA sites (Fig. 4b and Supplementary Fig. 15d). This difference was attributed partly to the two additional site types conserved in nematodes and partly to the higher fractions of hexamer and heptamer sites preferentially conserved in nematodes (Fig. 4c). *Drosophila*, which has intermediate 3′UTR lengths (median 224 nucleotides), had intermediate fractions of sites

conserved (Fig. 4c). Because the relative conservation of site types correlates well with their efficacy[18], species with shorter 3′UTRs presumably have increased relative efficacy of site types that impart marginal repression in vertebrates, such as most hexamer sites. With this increased miRNA targeting promiscuity, *C. elegans* could cope with shorter UTRs without sacrificing as much miRNA-mediated regulation.

MicroRNA sites were enriched in *C. elegans* 3′UTRs irrespective of conservation ($P < 10^{-3}$, binomial test; Fig. 4d). This enrichment was not observed in other regions of *C. elegans* mRNAs nor for human miRNA sites in any region of human mRNAs (Supplementary Fig. 15a). Perhaps in humans, the evolutionary depletion of detrimental miRNA sites balances the selective retention of beneficial sites, whereas in *C. elegans*, with its short UTRs, the depletion is not sufficient to balance the selective retention of beneficial sites (Supplementary Text). In this model, miRNA site enrichment would be a property of short 3′UTRs in any context. Indeed, enrichment of miRNA sites inversely correlated with mean 3′UTR length in both interspecies and intraspecies comparisons (Fig. 4d, e). Increased miRNA site density and increased efficacy of marginal site types in the context of short 3′UTRs are both likely to generalize to other *cis*-regulatory elements. Indeed, in *C. elegans*, the ~6,400 tandem UTR events occurred at one event per 560 nucleotides, a density over five times that reported in human UTRs[24,25].

3P-Seq provided a more comprehensive and reliable view of *C. elegans* 3′UTRs and the basis for insights into their formation, evolution, and regulation. The method should provide analogous results when applied to other eukaryotes with poorly annotated 3′UTRs, that is, most sequenced eukaryotes. 3P-Seq should also be informative for human studies, where it could shed light on shorter UTR isoforms, including those associated with cell proliferation and oncogenic transformation[26–28].

METHODS SUMMARY

Nematodes were grown and RNA isolated as described[29]. 3P-Seq was performed as outlined in Fig. 1a. Reads that both mapped to a single locus in the genome and possessed ≥2 3′-terminal adenylates (≥1 untemplated) were carried forward as 3P tags. Tags were iteratively clustered into representative cleavage sites and bridged to transcript models with RNA-Seq data (accession SRA003622.7)[10]. Poly(A) signals were identified as hexamers with position-dependent enrichment similar to AAUAAA. Cleavage sites 5′ of terminal exons indicated ALEs. For each last exon, cleavage sites mapping between the stop codon and the 3′-most 3P-Seq-supported cleavage site indicated tandem isoforms. Conservation analysis was as described[18], except five UTR conservation bins were used for *D. melanogaster* and four were used for *C. elegans*, in order to compensate for the smaller total sequence space of 3′UTRs in these species. For comparisons between mammals, flies and nematodes, pairs of species were chosen such that each had a 3′UTR nucleotide-level divergence rate of ~0.55.

1. Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309,** 1514–1518 (2005).
2. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136,** 719–730 (2009).
3. Ahringer, J. & Kimble, J. Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3′ untranslated region. *Nature* **349,** 346–348 (1991).
4. Wightman, B., Burglin, T. R., Gatto, J., Arasu, P. & Ruvkun, G. Negative regulatory sequences in the *lin-14* 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. *Genes Dev.* **5,** 1813–1824 (1991).
5. Merritt, C., Rasoloson, D., Ko, D. & Seydoux, G. 3′ UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr. Biol.* **18,** 1476–1482 (2008).
6. Rogers, A. *et al.* WormBase 2007. *Nucleic Acids Res.* **36,** D612–D617 (2008).
7. Mangone, M., Macmenamin, P., Zegar, C., Piano, F. & Gunsalus, K. C. UTRome.org: a platform for 3′UTR biology in *C. elegans. Nucleic Acids Res.* **36,** D57–D62 (2008).
8. Mangone, M. *et al.* The landscape of *C. elegans* 3′UTRs. *Science* **329,** 432–435 (2010).
9. Nam, D. K. *et al.* Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl Acad. Sci. USA* **99,** 6152–6156 (2002).
10. Hillier, L. W. *et al.* Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans. Genome Res.* **19,** 657–666 (2009).
11. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35,** D61–D65 (2007).
12. Nunes, N. M., Li, W., Tian, B. & Furger, A. A functional human poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J.* **29,** 1523–1536 (2010).
13. Evans, D. *et al.* A complex containing CstF-64 and the SL2 snRNP connects mRNA 3′ end formation and *trans*-splicing in *C. elegans* operons. *Genes Dev.* **15,** 2562–2571 (2001).
14. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl Acad. Sci. USA* **99,** 8796–8801 (2002).
15. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans. Mol. Cell* **31,** 67–78 (2008).
16. Chiang, H. R. *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24,** 992–1009 (2010).
17. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83–86 (2007).
18. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).
19. Zisoulis, D. G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans. Nature Struct. Mol. Biol.* **17,** 173–179 (2010).
20. Clark, A. M. *et al.* The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans. Nucleic Acids Res.* **38,** 3780–3793 (2010).
21. Lall, S. *et al.* A genome-wide map of conserved microRNA targets in *C. elegans. Curr. Biol.* **16,** 460–471 (2006).
22. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans. Nature* **403,** 901–906 (2000).
23. Abrahante, J. E. *et al.* The *Caenorhabditis elegans hunchback*-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. *Dev. Cell* **4,** 625–637 (2003).
24. Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33,** 201–212 (2005).
25. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
26. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320,** 1643–1647 (2008).
27. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA* **106,** 7028–7033 (2009).
28. Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138,** 673–684 (2009).
29. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans. Science* **294,** 858–862 (2001).
30. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3′-end processing. *Cell Mol. Life Sci.* **65,** 1099–1122 (2008).

Author Contributions C.H.J. performed the experiments and computational analyses of 3P-Seq data. R.C.F. performed the computational analyses of miRNA targeting and motif conservation. J.G.R. performed the computational analyses of miRNAs. All authors contributed to study design and manuscript preparation.

Author Information 3P-Seq reads and 3P tags were deposited at the GEO as fastq and BED files, respectively (GSE24924). MicroRNA genes were deposited at miRBase (miR4805–miR4816). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.P.B. (dbartel@wi.mit.edu).

METHODS

3P-Seq libraries. Nematodes were grown and RNA isolated as described[29]. For each library, 30 μg total RNA was enriched for polyadenylated mRNA (Dynabeads Oligo(dT)$_{25}$, Invitrogen). Enriched RNAs were ligated to a 5′-phosphorylated, 3′-biotinylated oligonucleotide adaptor (p-agcguguagggcaccauGCACATAC-Biotin; lowercase, RNA; uppercase, DNA) using the splint DNA oligonucleotide (ATGGTGCCCTACACGCTTTTTTTTTT) and 1 U Rnl II RNA ligase (New England Biolabs) in a 20 μl reaction for 16 h, according to the manufacturer's instructions. After phenol extraction and precipitation, the RNA was partially digested with 3 U RNase T1 (biochemistry grade, Ambion) in a 100 μl reaction for 20 min at 22 °C, and 3′ fragments were captured with 100 μl streptavidin-coated beads (Dynabeads M-280 streptavidin, Invitrogen) in 400 μl B buffer (5 mM Tris-Cl, pH 7.5, 0.5 mM EDTA, 1 M NaCl) for 15 min rotating at room temperature (~22 °C). After one wash in B buffer, beads were washed twice in 400 μl W buffer (10 mM Tris-Cl, pH 7.5, 1 mM EDTA, 50 mM NaCl) at 50 °C, then equilibrated in reverse transcription buffer (Invitrogen). The reverse transcription primer (GTATGTGCATGGTGCCCTACACGCT) was annealed and then extended with dTTP as the only deoxynucleoside triphosphate, using 1 U reverse transcriptase (Superscript III, Invitrogen) in 25 μl for 20 min at 48 °C. Polyadenylated RNA fragments were released into solution by adding 1 U RNase H (Invitrogen) and digesting for 25 min at 37 °C. After precipitation, fragments were ligated to a pre-adenylated 3′ adaptor (AppAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT-C3spacer, synthesized as in ref. 29) with 10 U T4 RNL1 (NEB) in a 10 μl reaction for 2 h at 22 °C, and ligation products were gel-purified (excising 75–300 nucleotide RNAs) and prepared for Illumina sequencing with a protocol used for strand-specific mRNA-Seq[31]. Because sequencing started at the residues corresponding to the 3′ ends of the RNA fragments, which for the 3P tags were all adenylates, cluster definition was somewhat compromised, which lowered the yield of 3P tags. In experiments performed after those described here, we obtained higher yields of 3P tags when defining the clusters from images in the middle of the run by starting the Illumina base calling with the middle images and then reading the images from the first part of the run after the clusters had been defined. A detailed 3P-Seq protocol is available at web.wi.mit.edu/bartel/pub/protocols.html.

Distal mRNA cleavage sites. The reverse complements of the sequencing reads were considered candidate 3P tags. These candidate tags were aligned to the *C. elegans* genome (WS190) with Bowtie[32], using alignment parameters '-q --solexa-quals -5 3 -l 25 -n 1 -e 240 -m 1' to allow for the presence of untemplated nucleotides at their 3′ termini. Sequences that both mapped to a single genomic locus and possessed ≥2 3′-terminal adenylates, of which at least one was untemplated, were carried forward as 3P-Seq tags. The most 3′-terminal non-adenosine base of each tag was considered a candidate cleavage site. Genomic loci were then marked off, using a set of RefSeq transcripts with non-redundant 3′ UTRs, with each locus corresponding to the region between the annotated 5′ terminus of the transcript and the annotated 5′ terminus of the downstream gene on the same genomic strand. Candidate sites mapping to each locus were sorted from most abundant to least, with equally abundant sites ordered 3′-most first. Clusters were then built from all sites within a 21-nucleotide window centred on the site with the most tags (combining data from all libraries). This process was iterated until all 3P tags were assigned to clusters (with some clusters containing only one tag). The central site of each cluster was then evaluated as a potential mRNA cleavage site using RNA-Seq data (accession number SRA003622.7)[10]. The number of RNA-Seq reads covering each base of the transcript was tallied, and a 50-nucleotide window was slid from the stop codon to the candidate terminus, after masking bases contained within annotated introns. A candidate site was assigned to an upstream protein-coding region if the median per-base RNA-Seq coverage in all windows was above 0. These sites were filtered further, requiring that the median per-base RNA-Seq coverage in the implied UTR was ≥5% that of the corresponding protein-coding region and that the maximum per-base RNA-Seq coverage in the UTR did not exceed five times that of the coding-region maximum. Among the sites that passed these filters, the distal cleavage site of a gene within the locus was the site of the 3′-most cluster that contributed ≥1% of the 3P tags from the locus.

Poly(A) signals. Genes with single UTRs were used to search for position-dependent enrichment of hexamer motifs near the cleavage site. To establish the region where PASs were expected to occur, AAUAAA enrichment was analysed at each position within the 50 nucleotides upstream of cleavage sites. At each position, significance was determined by the binomial test against the first-order Markov expectation for AAUAAA (Supplmentary Fig. 3). The region with significant AAUAAA enrichment (9–25 nucleotides upstream of the cleavage site) was analysed for enrichment of other hexamers, after removing the UTRs with AAUAAA in the region. The most significantly enriched alternative hexamer was identified as above, and sequences containing this hexamer were removed and the process was iterated another 13 times. Enrichment analysis was also performed by an alternative process, in which the first-order Markov expectations were replaced with the hexamer frequencies in an equally wide control window starting 50 nucleotides upstream of the cleavage site, and significant enrichment was determined by the Fisher's exact test. PASs were assigned to each cleavage site by searching the region 9–25 nucleotides upstream of the site, considering the 15 most significantly enriched hexamers (Supplementary Fig. 3 and Supplementary Table 4) and in cases of matches to more than one hexamer, giving preference to the one most significantly enriched in the global analysis that calculated enrichment using upstream control sequences.

Proximal alternative sites. For each Entrez gene, candidate cleavage sites (as defined above) were considered as proximal alternative cleavage sites if they (1) mapped between the 5′-most end and the 3′-most 3P-supported cleavage site of the gene, (2) were from clusters containing ≥1% of the tags from the gene, and (3) were from clusters containing two independent 3P-tags. Tags were considered independent if they either (1) were sequenced in independent libraries, (2) mapped to different cleavage sites, or (3) mapped to the same cleavage site but had different numbers of terminal adenylates. For each RefSeq transcript, proximal alternative sites 3′ of the stop codon were classified as proximal tandem sites. Candidate ALEs were identified by proximal alternative sites that mapped internally to genes, excluding the exons with any nucleotides 3′ of the distal-most stop codon for each Entrez gene annotation[33]. Identification of ALEs was particularly challenging in *C. elegans* for two reasons. First, many gene annotations had limited experimental validation. Second, *C. elegans* has a high density of genes[34], at least 15% of which are organized as operons[35]. Identification of ALEs thereby depended on experimental validation of the exons as alternative, that is, validation that the cleavage site is sometimes removed due to alternative splicing. ALEs were required to have the support of one of the following: (1) an EST omitting the cleavage site and aligning to both an upstream and downstream exon relative to the cleavage site; (2) an RNA-Seq read spanning the exon junction between the upstream and downstream exons; (3) 3P tags mapping to exons downstream of the ALE. In addition, ALEs were required to have an in-frame stop codon before the cleavage site. If no appropriate stop codon was annotated for novel ALEs, the nearest upstream stop codon was extended to the ALE cleavage site and the first in-frame stop codon was used.

Experimental evaluation of cleavage sites. Probes for RNase-protection experiments were designed to span proximal and distal cleavage sites of genes identified as having tandem UTRs either by both 3P-Seq and oligo(dT)-based methods (*rpl-12*, *kin-19*) or by only oligo(dT)-based methods (*ubc-18*)[8]. Templates for T7 transcription were amplified from N2 bristol genomic DNA using the following primer pairs: GAACAGCCCAATCCGTTGG, CAACACCAGTGTCTTTCGAT AC (*rpl-12*); CTCTTTTGGCTCCAAATGCC, AGGGTGTTACGGGAAATAGC (*kin-19*); GGAGCACACTCGAAAGCACG, CCGTGTTGTTATCGGCAACATC (*ubc-18*). Amplicons were cloned into a vector suitable for T7 transcription. Probes were body-labelled during *in vitro* transcription (MAXIscript, Ambion) and gel-purified on denaturing 5% acrylamide gels. RPAs were performed with 10^5 counts per minute probe and 15 μg total RNA, hybridized overnight at 42 °C and digested for 45 min in a 1:50 dilution of RNase A/T1 at 22 °C (RPA III, Ambion). Products were resolved on denaturing 5% acrylamide gels and visualized by phosphorimaging.

Conservation analyses. 3′ UTR alignments were extracted from Multi-Z alignments, (6-way for nematodes, 15-way for *Drosophila*) from the UCSC genome browser[36], starting with *D. melanogaster* RefSeq annotations for *Drosophila* UTRs. Conservation analyses were as described[18], except that five UTR conservation bins were used for *D. melanogaster* and four were used for *C. elegans*, to compensate for the smaller total sequence space of 3′UTRs in these species. Analyses of miRNA sites considered 60 *C. elegans* miRNA families with nucleotides 2–8 conserved throughout the *Caenorhabditis* clade (Supplementary Table 7) and 51 *Drosophila* miRNA families with nucleotides 2–8 conserved to *Drosophila pseudoobscura* (Supplementary Table 10). For imperfect site types, only the position of the bulged or mismatched nucleotide needed to be conserved, not the nucleotide itself. For comparisons between mammals, flies and nematodes (Fig. 4; Supplementary Fig. 15) pairs of species analysed were *Homo sapiens* and *Monodelphis domestica*, *Drosophila melanogaster* and *Drosophila willistoni*, and *Caenorhabditis elegans* and *Caenorhabditis briggsae*, each of which had a 3′UTR nucleotide-level divergence rate of ~0.55.

k-mer enrichment. For each type of miRNA site, 1,000 cohorts of control k-mers were chosen to match the site length, number of G+C nucleotides, and number of CpG dinucleotides. Enrichment was calculated by comparing the number of site occurrences in the region of interest to the mean number of occurrences for the controls. For each k-mer, the expected occurrence was determined using a first-order Markov model, and the P value was the fraction of control cohorts with ratios of observed-to-expected occurrences more extreme than that of the sites. When analysing enrichment in *Drosophila* and human 3′UTRs (Fig. 4 and Supplementary Fig. 15), only RefSeq annotations with 'validated' status were used.

31. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).

32. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
33. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* **35,** D26–D31 (2007).
34. The Celegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282,** 2012–2018 (1998).
35. Blumenthal, T. Trans-splicing and operons. *WormBook* **25,** 1–9 (2005).
36. Karolchik, D. *et al.* The UCSC genome browser database. *Nucleic Acids Res.* **31,** 51–54 (2003).

LETTER

Telomerase reactivation reverses tissue degeneration in aged telomerase–deficient mice

Mariela Jaskelioff[1], Florian L. Muller[1], Ji–Hye Paik[1], Emily Thomas[1], Shan Jiang[1], Andrew C. Adams[2], Ergun Sahin[1], Maria Kost-Alimova[1], Alexei Protopopov[1], Juan Cadiñanos[1], James W. Horner[1], Eleftheria Maratos–Flier[2] & Ronald A. DePinho[1]

An ageing world population has fuelled interest in regenerative remedies that may stem declining organ function and maintain fitness. Unanswered is whether elimination of intrinsic instigators driving age-associated degeneration can reverse, as opposed to simply arrest, various afflictions of the aged. Such instigators include progressively damaged genomes. Telomerase-deficient mice have served as a model system to study the adverse cellular and organismal consequences of wide-spread endogenous DNA damage signalling activation *in vivo*[1]. Telomere loss and uncapping provokes progressive tissue atrophy, stem cell depletion, organ system failure and impaired tissue injury responses[1]. Here, we sought to determine whether entrenched multi-system degeneration in adult mice with severe telomere dysfunction can be halted or possibly reversed by reactivation of endogenous telomerase activity. To this end, we engineered a knock-in allele encoding a 4-hydroxytamoxifen (4-OHT)-inducible telomerase reverse transcriptase-oestrogen receptor (TERT-ER) under transcriptional control of the endogenous TERT promoter. Homozygous TERT-ER mice have short dysfunctional telomeres and sustain increased DNA damage signalling and classical degenerative phenotypes upon successive generational matings and advancing age. Telomerase reactivation in such late generation TERT-ER mice extends telomeres, reduces DNA damage signalling and associated cellular checkpoint responses, allows resumption of proliferation in quiescent cultures, and eliminates degenerative phenotypes across multiple organs including testes, spleens and intestines. Notably, somatic telomerase reactivation reversed neurodegeneration with restoration of proliferating Sox2[+] neural progenitors, Dcx[+] newborn neurons, and Olig2[+] oligodendrocyte populations. Consistent with the integral role of subventricular zone neural progenitors in generation and maintenance of olfactory bulb interneurons[2], this wave of telomerase-dependent neurogenesis resulted in alleviation of hyposmia and recovery of innate olfactory avoidance responses. Accumulating evidence implicating telomere damage as a driver of age-associated organ decline and disease risk[1,3] and the marked reversal of systemic degenerative phenotypes in adult mice observed here support the development of regenerative strategies designed to restore telomere integrity.

Accelerating structural and functional decline across diverse organ systems is observed in the aged[1,3,4]. The loss of genome integrity and associated DNA damage signalling and cellular checkpoint responses are well-established intrinsic instigators that drive tissue degeneration during ageing[5]. Of particular relevance to this study, age-progressive loss of telomere function in mice has been shown to provoke widespread p53 activation resulting in activation of cellular checkpoints of apoptosis, impaired proliferation and senescence, compromised tissue stem cell and progenitor function, marked tissue atrophy and physiological impairment in many organ systems[1,6].

Mounting evidence in humans has also provided strong association of limiting telomeres with increased risk of age-associated disease[7] and with onset of tissue atrophy and organ system failure in degenerative diseases such as ataxia-telangiectasia, Werner syndrome, dyskeratosis congenita and liver cirrhosis, among others[1,3]. In cell-based models of ataxia-telangiectasia and Werner syndrome, enforced TERT can restore normal cellular proliferative potential[8]. These findings build on seminal cell culture studies showing that enforced TERT expression can endow primary human cells with unlimited replicative potential[9]. Importantly, TERT overexpression in epithelial tissues of cancer-resistant mice leads to extended median lifespan[10]. In addition, intercrossing wild-type and late generation *mTerc*[−/−] mice with severe degenerative phenotypes results in healthy offspring[11], indicating that viable late generation *mTerc*[−/−] germ cells can be restored to normal telomere function on introduction of a wild-type *mTerc* allele at the time of fertilization. However, to our knowledge, there are no genetic or pharmacological studies showing somatic reversal of age-related degenerative phenotypes driven by endogenous genotoxic stresses in adult mammals. Here, in telomerase-deficient mice experiencing severe tissue degeneration, we investigated whether endogenous telomerase-mediated restoration of telomere function throughout the organism would quell DNA damage signalling and either arrest, or possibly reverse, cellular checkpoint responses and associated tissue atrophy and dysfunction. Notably, the mice enlisted into this study are adults exhibiting significant progeroid phenotypes.

Construction and functional validation of the germline TERT-ER knock-in allele are detailed in Supplementary Fig. 1. In the absence of 4-OHT, ER fusion proteins remain in an inactive misfolded state[12] and thus we first sought to verify whether mice homozygous for TERT-ER recapitulated the classical premature ageing phenotypes of mice null for *mTerc* or *mTert*. To that end, mice heterozygous for TERT-ER (hereafter G0[TERT-ER]) were intercrossed to produce first generation mice homozygous for TERT-ER (G1[TERT-ER]) which were then intercrossed to produce successive G2, G3 and G4[TERT-ER] cohorts. G1–G4[TERT-ER] cells have no detectable telomerase activity (Fig. 1a). Accordingly, G4[TERT-ER] primary splenocytes had hallmark features of short dysfunctional telomeres, including decreased telomere-specific fluorescence *in situ* hybridization (FISH) signal and Robertsonian fusions (Fig. 1b, e, f). Moreover, G4[TERT-ER] fibroblasts failed to divide after five to six passages and adopted a flat, senescence-like morphology (Fig. 1c, d). Adult G4[TERT-ER] mice showed widespread tissue atrophy, particularly in highly proliferative organs including extreme testicular atrophy and reduced testes size due to apoptotic elimination of germ cells, resulting in decreased fecundity (Fig. 2a, d and Supplementary Fig. 2a), marked splenic atrophy with accompanying increased 53BP1 (also known as Trp53bp1) foci consistent with DNA damage (Fig. 2b, e, h) and intestinal crypt depletion and villus atrophy in conjunction with numerous apoptotic crypt cells and increased 53BP1 foci (Fig. 2c, f, i and Supplementary Fig. 2b). Finally, median survival of G4[TERT-ER] mice is significantly decreased relative to that of telomere intact mice (43.5 versus 86.8 weeks, ***$P < 0.0001$, Supplementary Fig. 2f). Thus, G4[TERT-ER] mice phenocopy late generation *mTert*[−/−] and *mTerc*[−/−] animals[13,14], indicating that TERT-ER is inactive in the absence of 4-OHT.

[1]Belfer Institute for Applied Cancer Science and Departments of Medical Oncology, Medicine and Genetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Division of Endocrinology, Diabetes & Metabolism, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA.
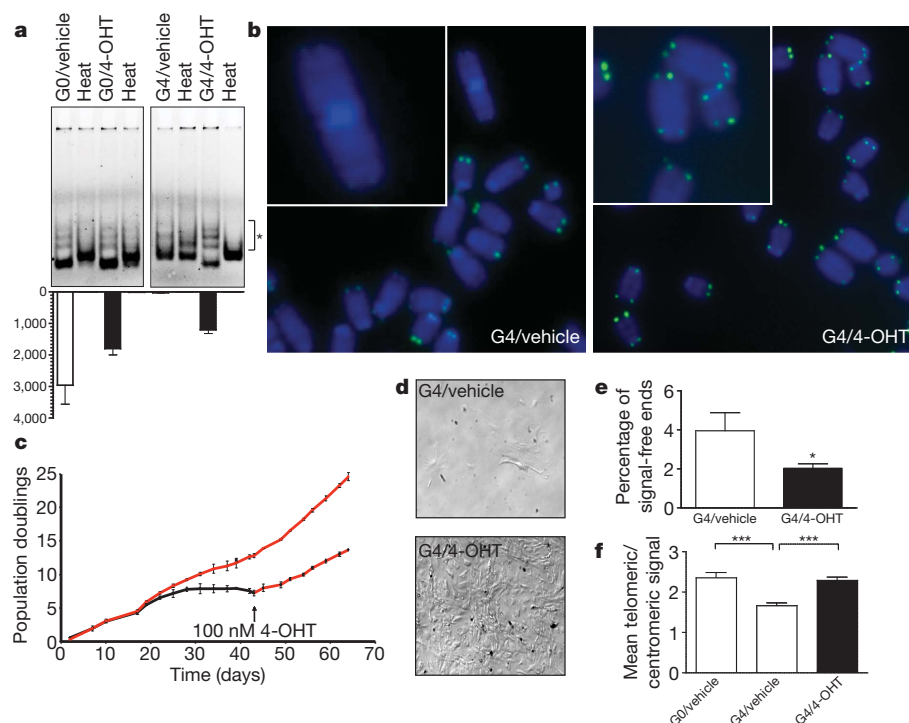
Figure 1 | 4-OHT-dependent induction of telomerase activity in TERT-ER cells. **a**, Telomerase activity in eNSCs (*, telomerase products) (top); real-time quantification of reactions above (bottom). **b**, Representative G4$^{TERT-ER}$ splenocyte metaphases. **c**, Proliferation of adult G4$^{TERT-ER}$ fibroblasts ($n = 3$) in media with vehicle (black) or 4-OHT (red). **d**, Representative image of G4$^{TERT-ER}$ fibroblasts (passage 6) in media with 4-OHT (bottom) or vehicle (top). **e**, Signal-free ends in primary splenocyte metaphases, 15 metaphases per sample, $n = 2$ (*$P < 0.05$). **f**, Mean telomere-FISH signal in primary splenocyte interphases, normalized to centromeric signal, $n = 3$ (***$P < 0.0001$). Open bars correspond to vehicle-treated and filled bars to 4-OHT-treated, error bars represent s.d.

Next, we assessed the impact of telomerase reactivation on telomere dysfunction-induced proliferative arrest. On passage of adult G4$^{TERT-ER}$ fibroblast cultures, cells adopted flat senescent-like morphology at approximately five population doublings (Fig. 1d, upper panel). These quiescent cultures showed prominent G0/G1 accumulation in the cell cycle by fluorescence-activated cell sorting (FACS) analysis and rare cell division events by time-lapse video microscopy (not shown). However, upon replating these cells in media containing 100 nM 4-OHT, telomerase reactivation led to elongated telomeres, prompt resumption of proliferation over greater than eight additional passages tested, and reduction in the G0/G1 phase fraction (Fig. 1c; data not shown). Coincidently, high levels of cyclin-dependent kinase inhibitor, p21^{CIP1} (also known as Cdkn1a), declined upon 4-OHT treatment of the G4$^{TERT-ER}$ cultures, allowing cell cycle re-entry (Supplementary Fig. 2e). This pattern of p21^{CIP1} regulation aligns with previous work documenting its role as a key mediator of cell cycle arrest induced by telomere dysfunction in mouse tissues[15]. Parallel G0 or G4$^{TERT-ER}$ fibroblasts maintained in 4-OHT at initial isolation did not undergo passage-induced senescence and instead showed sustained proliferation (>20 passages; Fig. 1c, d).

These cell-based studies prompted systemic analyses of the impact of 4-OHT-mediated telomerase reactivation in the setting of entrenched tissue degeneration. At the end of 4 weeks of continuous 4-OHT exposure, documentation of telomerase-mediated telomere restoration and function in G4$^{TERT-ER}$ tissues included increased telomere-FISH signal in primary splenocytes (Fig. 1b, e, f), decreased p53 activation and expression of p21^{CIP1} in liver (Supplementary Fig. 2d, e), and marked decrease in 53BP1 foci in splenocytes (Fig. 2b, e) and intestinal crypt cells (Fig. 2c, f). These molecular changes paralleled striking tissue rejuvenation including reduced apoptosis of testes germ cells (data not shown) and intestinal crypt cells (Supplementary Fig. 2b, i), reduced tissue atrophy with restoration in normal testes and spleen size (Fig. 2d, h and Supplementary Fig. 2a) and, most strikingly, increased fecundity (Fig. 2g). Moreover, median survival increased in G4$^{TERT-ER}$ mice treated with a 4-week course of 4-OHT (**$P < 0.005$, Supplementary Fig. 2f). Sustained 4-OHT treatment had no effect on G0$^{TERT-ER}$ age- and gender-matched controls which were included in all experiments. Together, these data indicate that, despite an entrenched degenerative state, endogenous telomerase reactivation

results in marked extinction of DNA damage signalling, alleviation of cellular checkpoint responses and reversal of tissue atrophy in highly proliferative organ systems of the late generation TERT-ER mice.

Although the marked impact of telomerase reactivation on highly proliferative organs is encouraging, we sought to assess more intensively the potential benefits on brain health, which is a prime determinant of age-progressive declining health in humans. Along these lines, it is worth noting that the ageing mammalian brain shows accumulating DNA damage[16] and a progressive restriction of neurogenesis and impaired re-myelination due to a decline in neural stem and progenitor cell proliferation and differentiation[17]. As neural stem/progenitor cells (hereafter NSCs) support neurogenesis, particularly in the subventricular zone (SVZ), we first examined the properties of NSCs derived from adult G0 and G4$^{TERT-ER}$ mice. As reported previously for late generation $mTerc^{-/-}$ mice[6,14,18], vehicle-treated G4$^{TERT-ER}$ NSC cultures showed decreased self-renewal activity relative to G0$^{TERT-ER}$ controls and this defect was partially corrected with 4-OHT treatment (Fig. 3a, d). G4$^{TERT-ER}$ neurospheres were not only rarer but also smaller in diameter than G0$^{TERT-ER}$ controls, and their average diameter was restored to normal by 4-OHT treatment (Fig. 3a and Supplementary Fig. 2c). These self-renewal profiles tracked with activated p53-mediated DNA damage signalling in vehicle-treated G4$^{TERT-ER}$ NSC cultures, which was extinguished with 4-OHT treatment and absent in the G0$^{TERT-ER}$ controls (Fig. 3b, e). Examination of NSC differentiation capacity revealed significant (twofold) reduction in G4$^{TERT-ER}$ NSC capacity to generate neurons relative to 4-OHT-treated G4$^{TERT-ER}$ cultures and 4-OHT- or vehicle-treated G0$^{TERT-ER}$ controls (Fig. 3c, f). Consistent with previous work[14,18], there was no impact on astrocyte differentiation (data not shown).

On the basis of these cell culture observations, we examined the SVZ, a region where NSCs reside and have an active role in adult brain physiology. In adult mice, NSCs give rise to transit-amplifying progenitor cells that divide rapidly and contribute to generation of neuroblasts, astrocytes and myelinating oligodendrocytes. Consistent with previous reports of an SVZ proliferation defect in $mTerc^{-/-}$ mice[6,14,18] and wild-type aged mice[19], vehicle-treated G4$^{TERT-ER}$ mice show a profound decrease in proliferating (Ki67$^+$) cells in the SVZ relative to G0$^{TERT-ER}$ controls. Notably, 4-OHT-treated G4$^{TERT-ER}$ mice show a striking, albeit partial, restoration of proliferation following only
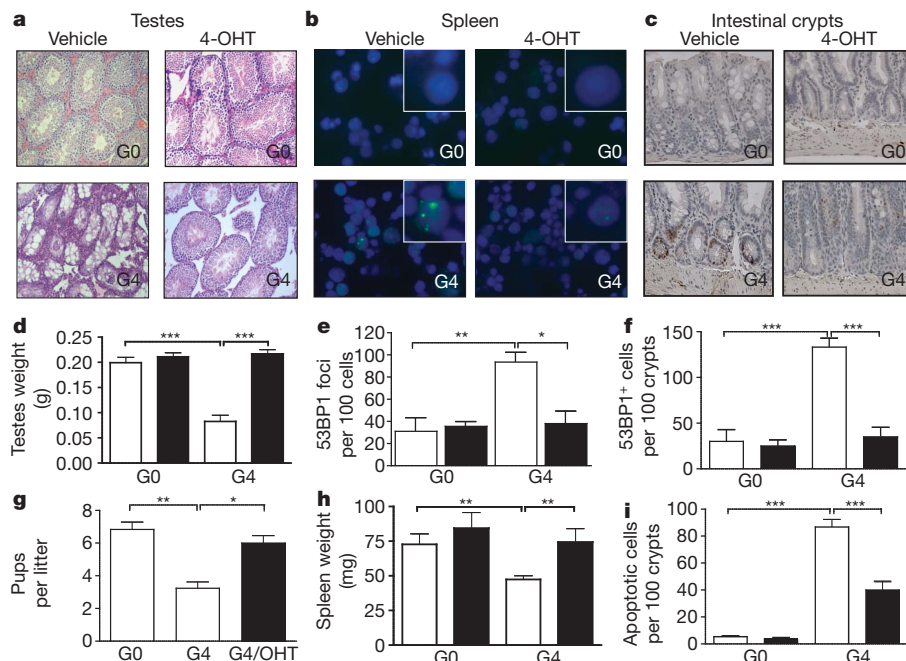
Figure 2 | Telomerase activation in adult TERT-ER mice. a–c, Representative images of tissues from experimental and control mice. **a,** Haematoxylin and eosin-stained sections of testes. **b,** Primary splenocytes stained for 53BP1. **c,** Small intestine sections stained for 53BP1. **d,** Testes weight of adult males (30–50-week-old, $n \geq 10$). **e,** 53BP1 nuclear foci per 100 nuclei ($n = 3$). **f,** 53BP1 nuclear foci per 100 crypts ($n = 4$). **g,** Litter sizes ($n = 3$); **h,** Spleen weights ($n \geq 6$). **i,** Apoptotic cells per 100 intestinal crypts ($n \geq 20$). ***$P = 0.0001$, **$P < 0.005$, *$P < 0.05$. Open bars correspond to vehicle-treated and filled bars to 4-OHT-treated groups, error bars represent s.d.

4 weeks of treatment (Fig. 4, first row). This resumed SVZ proliferation mirrors well restoration of Sox2$^+$ cells, a marker of NSCs (Fig. 4, second row), and doublecortin (Dcx)-positive cells, an early neuronal lineage marker, together demonstrating preservation of neural stem/progenitor reserves and their neurogenic capacity *in vivo* (Fig. 4, third row). Finally, quantitative FISH analysis shows telomere elongation in the SVZ after 4 weeks of 4-OHT treatment (Supplementary Fig. 3). Thus, the markedly constrained neural progenitor proliferation and neurogenesis profile associated with telomere dysfunction can be ameliorated by reactivation of endogenous telomerase activity.

To test the hypothesis that telomerase reactivation leads to tissue rejuvenation, we conducted detailed morphological and functional fitness analyses of different brain structures upon telomerase reactivation. First, we examined the white matter of the corpus callosum and observed that aged G4$^{TERT-ER}$ mice have far fewer Olig2$^+$ mature oligodendrocytes (Fig. 4, fourth row). This cellular deficiency is associated with reduced brain weight (Fig. 5a, b) and significantly thinner myelin sheathing of neurons with g ratios (numerical ratio between the diameter of the axon proper and the outer diameter of the myelinated fibre) of 0.7756 ± 0.0054 for G4$^{TERT-ER}$ mice versus 0.7032 ± 0.0049 for G0$^{TERT-ER}$ (mean ± s.e.m., ***$P < 0.0001$) (Fig. 5c, d). Remarkably, endogenous telomerase reactivation reinstates normal numbers of mature oligodendrocytes (Fig. 4) and reverses the hypomyelination

phenotype at the level of mean myelin sheath diameters (with g ratios of 0.7058 ± 0.0006 and 0.7164 ± 0.0063 for 4-OHT-treated G4 and G0$^{TERT-ER}$ mice, respectively) (Fig. 5c, d). Furthermore, a 4-OHT treatment course of only 4 weeks is sufficient to cause significant partial reversion of the brain size defect, with G4$^{TERT-ER}$ brain weights increasing from $77.3 \pm 3.3\%$ of G0$^{TERT-ER}$ brain weights in the vehicle group to $89.7 \pm 4.0\%$ in the 4-OHT group (Fig. 5a, b). Importantly, telomere elongation can be detected in the corpus callosum after 4 weeks of telomerase reactivation (Supplementary Fig. 3c). Thus, endogenous telomerase reactivation exerts a swift impact on oligodendrocyte proliferation and differentiation, and promotes repopulation of white matter structures with mature oligodendrocytes and active myelin deposition.

Lastly, we investigated the physiological effect of telomere dysfunction and telomerase reactivation on olfactory function. Age-associated hyposmia, as evidenced by an increased olfactory threshold and a reduced ability in odour identification and discrimination, is a well established phenomenon in aged humans[20]. In rodents, ageing is associated with diminished olfactory neurogenesis and deficits in fine olfactory discrimination[19,21]. Olfactory interneurons in the olfactory bulb that receive and process information from the olfactory sensory neurons in the olfactory epithelium derive from SVZ stem cells[2]. Rodents demonstrate avoidance responses towards predators' odorants as well as spoiled smells like aliphatic acids, aliphatic aldehydes and alkyl amines, which are
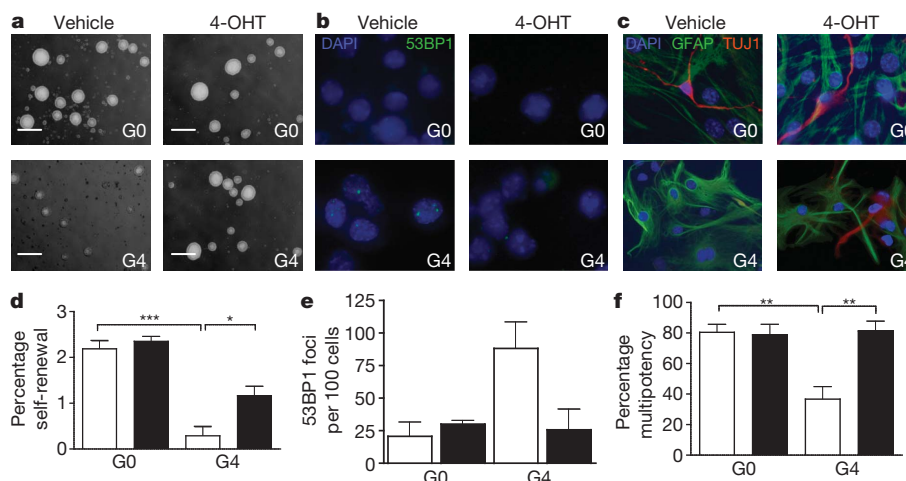

Figure 3 | Neural stem cell function following telomerase reactivation *in vitro*. **a–c,** Representative images of experimental and control mice-derived NSCs. **a,** Secondary neurospheres. **b,** Differentiated NSCs stained with 53BP1 or **c,** GFAP and TUJ1 antibodies. **d,** Self-renewal capacity of secondary neurospheres ($n = 4$) ***$P < 0.0001$, *$P < 0.001$. **e,** 53BP1 nuclear foci per 100 cells (>400 nuclei per culture, $n = 3$). **f,** Multipotency (GFAP$^+$/TUJ1$^+$) of NSCs ($n = 4$; 308 wells per culture condition) **$P = 0.0066$. Scale bar, 100 μm. Open bars correspond to vehicle-treated and filled bars to 4-OHT-treated groups, error bars represent s.d.
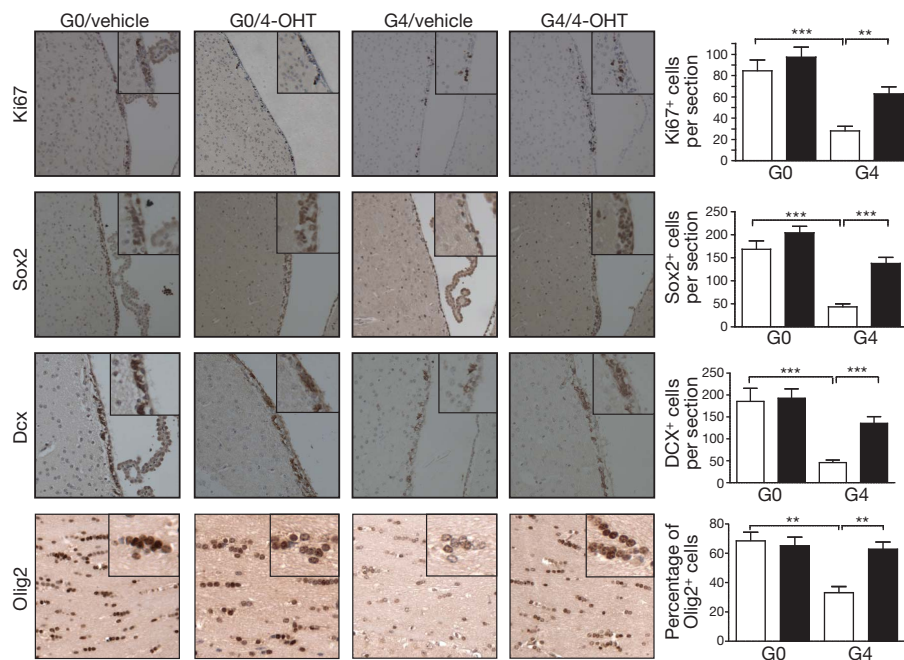
Figure 4 | NSC proliferation and differentiation following telomerase reactivation *in vivo*. NSC proliferation and neurogenesis were measured by Ki67, Sox-2 and Dcx expression in SVZ from experimental and control mice. Mature oligodendrocytes in the corpus callosum were stained with anti-Olig2 antibody. Equivalent coronal sections ($n > 10$) were scored in a blinded fashion by laser scanning and plotted on the right panels. ×20 (SVZ) or ×40 (corpus callosum) objectives were used. *** $P < 0.0001$, **$P = 0.0022$. Open bars correspond to vehicle-treated and filled bars to 4-OHT-treated groups, error bars represent s.d.

processed in the olfactory bulb[22]. Given the marked decrease in SVZ neurogenesis of G4[TERT-ER] mice and the fact that the olfactory bulb retains high telomerase activity in adult wild-type mouse brains[23], we sought to determine whether telomere dysfunction results in a functional deficit of these mice to detect and process odorants for elicitation of instinctive avoidance/defensive behaviours.

Pathology within the olfactory epithelium which may be considered a basis of age-related olfactory dysfunction, was ruled out by confirmation of grossly normal histology of the olfactory epithelium in both cohorts (Supplementary Fig. 4). Next, we ruled out alterations in exploration behaviour and overall locomotion by monitoring total distance travelled by the animals in the absence of odorants, which was similar for all experimental groups (Supplementary Table 1; Fig. 5e). We then performed innate avoidance tests using serially diluted 2-methylbutyric acid (2-MB), an odorant that rouses innate aversive responses in mice. Whereas G0[TERT-ER] mice demonstrated avoidance

responses at all 2-MB concentrations tested (1.87×10^{-4} M to 1.87×10^{-6} M), G4 mice showed attraction/neutral behaviours at concentrations lower than 1.87×10^{-4} M (Fig. 5e, f). Strikingly, following only 4 weeks of 4-OHT treatment, the performance of G4[TERT-ER] mice was markedly improved, with avoidance behaviours being apparent at all 2-MB concentrations (Fig. 5e, g). Accordingly, the frequency of entry into the odour zone was higher for vehicle-treated G4[TERT-ER] mice than for the other three experimental groups (Supplementary Table 2). These findings are consistent with significant alleviation of the olfactory defect stemming from the documented wave of telomerase-mediated SVZ neurogenesis and oligodendrocyte maturation which would promote repopulation of olfactory bulbs with functional interneurons and improve olfactory neuron function via remyelination.

Here, we report the generation of a novel mouse model to explore the impact of physiological telomerase reactivation across diverse adult cell types and organ systems. In G4[TERT-ER] mice with advanced degenerative
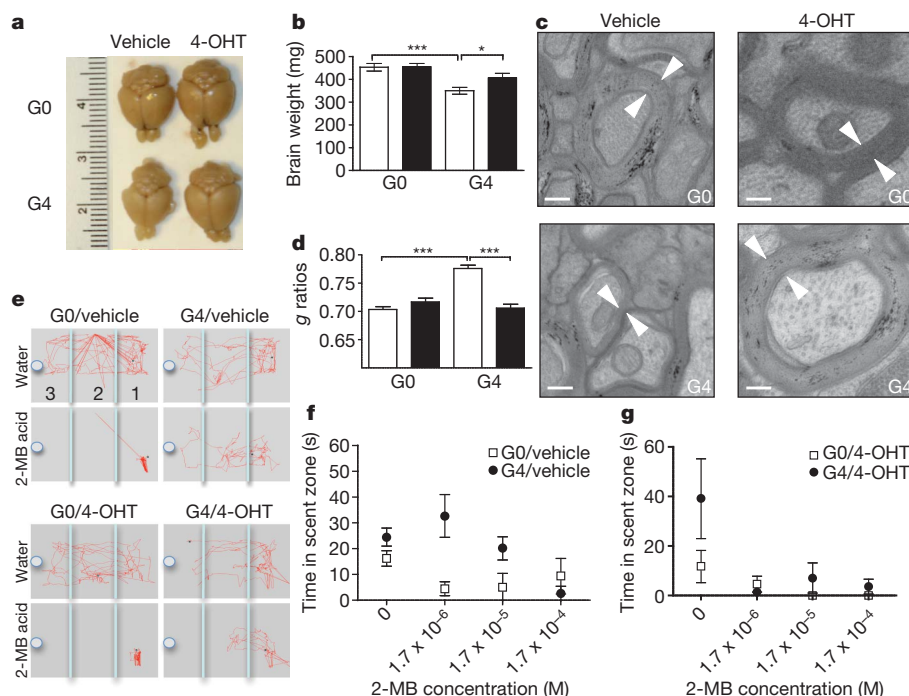

Figure 5 | Brain size, myelination, and olfactory function following telomerase reactivation. **a**, Representative brains from age-matched experimental and control animals. **b**, Brain weights, $n \geq 10$, ***$P = 0.0004$, *$P = 0.02$. **c**, Representative electron micrographs of myelinated axonal tracts in corpus callosum, arrow heads indicate myelin sheath width (×12,000). Scale bars, 200 nm. **d**, *g* ratios (inner/outer radii) ($n = 2$, >150 axons per mouse) ***$P < 0.0001$. **e**, Representative tracings of experimental and control mice during 3-min exposure to water or 2-MB. **f, g**, Time spent in scent zone 3 with water or 2-MB for vehicle- or 4-OHT-treated G0[TERT-ER] (squares) and G4[TERT-ER] (circles) mice; $n = 4$. Error bars represent s.d., except in (**d**) (s.e.m.).

phenotypes, short-term telomerase reactivation restored telomere reserves, quelled DNA damage signalling, and alleviated cellular checkpoint responses in several high-turnover organ systems with significant functional impact including increased fecundity. From this, we speculate that some tissue stem/progenitor cells are retained in a quiescent and intact state yet can be enlisted to resume normal repopulating function upon elimination of genotoxic stress at telomeres. Despite chromosomal instability, the brief course of telomerase reactivation was not sufficient to promote carcinogenesis (data not shown), a finding consistent with a role for telomerase in promoting progression of established neoplasms[24]. However, it remains possible that more prolonged telomerase reactivation schedules or applications in later life may provoke carcinogenesis.

As noted, age-associated compromise in mammalian brain function is associated with extensive accumulation of DNA damage and progressive reduction in neurogenesis and myelination. Indeed, many aspects of this central nervous system decline are accelerated and worsened in the setting of telomere dysfunction (refs 25, 26, this study). Our data establish that telomerase reactivation in adult mice with telomere dysfunction can restore SVZ neurogenesis and, consistent with its role in sustaining new olfactory bulb neurons, can ameliorate odour detection with improved performance in innate odour avoidance tests. These results are consistent with previous studies showing that prolonged inhibition of neurogenesis in the SVZ has a negative effect on odour detection thresholds[27]. In conclusion, this unprecedented reversal of age-related decline in the central nervous system and other organs vital to adult mammalian health justify exploration of telomere rejuvenation strategies for age-associated diseases, particularly those driven by accumulating genotoxic stress.

METHODS SUMMARY

TERT-ER mice were generated with traditional knock-in methods and following standard breeding protocol of successive generations of telomerase-deficient mice[13]. All studies were performed on adult males. 4-OHT time-release pellets (2.5 mg; Innovative Research of America) were inserted subcutaneously to reach steady state blood levels of $1\,\mathrm{ng\,ml^{-1}}$ 4-OHT. For neurosphere assays, SVZs were dissected, dispersed into a single-cell suspension and plated in neurobasal media supplemented with EGF, bFGF and 100 nM 4-OHT or vehicle. For multipotentiality assays, neurospheres were transferred to differentiation media (1% FBS). For histological studies, mice were perfused with 10% formalin; equivalent coronal sections were stained with indicated antibodies following standard immunohistochemistry protocol. Laser scanning cytometric quantification was performed with an iCys Research Imaging Cytometer (Compucyte). For innate olfactory avoidance tests, mice were fasted for 20 h and habituated for 20 min to the test cage where their responses were recorded on a video camera mounted above the test chamber. A filter paper scented with water or progressively higher concentrations of 2-methylbutyric acid was placed in the cage and mouse behaviour was recorded for 3 min. NoldusEthovision v3.1 behavioural analysis software was used to determine innate avoidance behaviour (time spent in the third of the cage containing the scented filter paper).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Sahin, E. & Depinho, R. A. Linking functional decline of telomeres, mitochondria and stem cells during ageing. *Nature* **464,** 520–528 (2010).
2. Whitman, M. C. & Greer, C. A. Adult neurogenesis and the olfactory system. *Prog. Neurobiol.* **89,** 162–175 (2009).
3. Sharpless, N. E. & DePinho, R. A. How stem cells age and why this makes us grow old. *Nature Rev. Mol. Cell Biol.* **8,** 703–713 (2007).
4. Ju, Z. & Lenhard Rudolph, K. Telomere dysfunction and stem cell ageing. *Biochimie* **90,** 24–32 (2008).
5. Hoeijmakers, J. H. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361,** 1475–1485 (2009).
6. Ferrón, S. *et al.* Telomere shortening and chromosomal instability abrogates proliferation of adult but not embryonic neural stem cells. *Development* **131,** 4059–4070 (2004).
7. Cawthon, R. M. *et al.* Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* **361,** 393–395 (2003).
8. Wyllie, F. S. *et al.* Telomerase prevents the accelerated cell ageing of Werner syndrome fibroblasts. *Nature Genet.* **24,** 16–17 (2000).
9. Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279,** 349–352 (1998).
10. Tomás-Loba, A. *et al.* Telomerase reverse transcriptase delays aging in cancer-resistant mice. *Cell* **135,** 609–622 (2008).
11. Samper, E., Flores, J. M. & Blasco, M. A. Restoration of telomerase activity rescues chromosomal instability and premature aging in $Terc^{-/-}$ mice with short telomeres. *EMBO Rep.* **2,** 800–807 (2001).
12. Metzger, D. *et al.* Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc. Natl Acad. Sci. USA* **92,** 6991–6995 (1995).
13. Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91,** 25–34 (1997).
14. Wong, K. K. *et al.* Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. *Nature* **421,** 643–648 (2003).
15. Choudhury, A. R. *et al. Cdkn1a* deletion improves stem cell function and lifespan of mice with dysfunctional telomeres without accelerating cancer formation. *Nature Genet.* **39,** 99–105 (2006).
16. Best, B. P. Nuclear DNA damage as a direct cause of aging. *Rejuvenation Res.* **12,** 199–208 (2009).
17. Drapeau, E. & Nora Abrous, D. Stem cell review series: role of neurogenesis in age-related memory disorders. *Aging Cell* **7,** 569–589 (2008).
18. Ferrón, S. R. *et al.* Telomere shortening in neural stem cells disrupts neuronal differentiation and neuritogenesis. *J. Neurosci.* **29,** 14394–14407 (2009).
19. Enwere, E. *et al.* Aging results in reduced epidermal growth factor receptor signaling, diminished olfactory neurogenesis, and deficits in fine olfactory discrimination. *J. Neurosci.* **24,** 8354–8365 (2004).
20. Lafreniere, D. & Mann, N. Anosmia: loss of smell in the elderly. *Otolaryngol. Clin. North Am.* **42,** 123–131 (2009).
21. Ma, D. K. *et al.* Activity-dependent extrinsic regulation of adult olfactory bulb and hippocampal neurogenesis. *Ann. NY Acad. Sci.* **1170,** 664–673 (2009).
22. Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450,** 503–508 (2007).
23. Caporaso, G. L. *et al.* Telomerase activity in the subventricular zone of adult mice. *Mol. Cell. Neurosci.* **23,** 693–702 (2003).
24. Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31,** 9–18 (2010).
25. Zhang, P., Dilley, C. & Mattson, M. P. DNA damage responses in neural cells: focus on the telomere. *Neuroscience* **145,** 1439–1448 (2007).
26. Lee, J. *et al.* Telomerase deficiency affects normal brain functions in mice. *Neurochem. Res.* **35,** 211–218 (2010).
27. Breton-Provencher, V. *et al.* Interneurons produced in adulthood are required for the normal functioning of the olfactory bulb network and for the execution of selected olfactory behaviors. *J. Neurosci.* **29,** 15245–15257 (2009).

Author Contributions M.J. and R.A.D. designed and guided the research; M.J., F.L.M., J.-H.P., E.S., E.T., S.J. and M.K.-A. performed research. J.C. and J.W.H. generated the TERT-ER mouse. M.J., F.L.M., A.C.A., A.P., E.M.-F. and R.A.D. analysed data. M.J. and R.A.D. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.D. (ron_depinho@dfci.harvard.edu).

METHODS

Generation of TERT-ER mice. A knock-in targeting vector containing the ERT2-LBD domain upstream and in frame with the *mTert* genomic sequence (exon 1 through intron 2) and a *Lox-pgk-Neo-Lox* fragment was introduced into ES cells. Neomycin-resistant clones yielded five independent lines, two of which were injected into C57BL/6 blastocysts and implanted into surrogate mothers, yielding 10 high-percentage chimaeras. Germline transmission was confirmed by crossing the chimaeras to C57BL/6 females. Heterozygous TERT-ERneo animals were crossed to EIIa-Cre animals to delete the NeoR cassette and further intercrossed to homozygosity. The EIIa-Cre allele was then bred out of the line and heterozygous animals were backcrossed to C57BL/6 at least three times. From this point, standard breeding protocol of successive generations of telomerase-deficient mice was followed[13,28]. All studies were performed on adult (30–35-week-old) males, heterozygous (G0$^{TERT-ER}$) or homozygous (G4$^{TERT-ER}$) for this allele, unless otherwise noted. 4-OHT time-release pellets (2.5 mg; Innovative Research of America) were inserted subcutaneously to reach steady state blood levels of 1 ng ml^{-1} 4-OHT. Mice were maintained in specific pathogen-free (SPF) conditions at Dana-Farber Cancer Institute. All manipulations were performed with IACUC approval.

Histology and electron microscopy. Brains from animals perfused with 10% formalin were further fixed for 24 h and coronally sectioned using a brain matrix (Electron Microscopy Sciences). Equivalent sections were used for chromogenic immunohistochemistry, which was performed according to standard procedures. Antibodies used include anti-Ki67 (Dako), anti-53BP1 (Bethyl Labs), anti-Sox-2 and anti-Dcx (Santa Cruz Biotechnology) and anti-Olig-2 (Chemicon). For immunofluorescence studies, cells were fixed in 4% paraformaldehyde (PFA) in phosphate-buffered saline for 10 min, permeabilized (50 mM NaCl, 3 mM MgCl$_2$, 200 mM sucrose, 10 mM HEPES pH 7.9, 0.5% TX-100) for 5 min, and then stained with primary antibodies against 53BP1 (Bethyl Labs), TUJ1 (Chemicon), GFAP (Dako) and secondary antibodies conjugated to Alexa Fluor-488 or Alexa Fluor-568 (Molecular Probes). Cells were mounted in DAPI-containing antifade solution (Vector). Foci were scored by eye from a minimum of 300 randomly chosen nuclei by using a ×40 objective, and scoring was performed in a blinded manner with respect to genotype. Immunofluorescence images were captured in greyscale for each fluorophore and were merged by compilation in respective red-green-blue (RGB) channels using Adobe Photoshop CS 8.0. For apoptosis assays, sections from paraffin-embedded testes were deparaffinized and processed for apoptotic staining (terminal deoxynucleotidyl transferase-mediated dUTP-biotin nick end labelling, TUNEL) according to the manufacturer's instructions (Chemicon). For electron microscopy studies, animals were perfused for 30 min in Karnovsky's solution B, brains were further fixed for 24 h and delivered to the Harvard Medical School Electron Microscopy Facility for embedding, sectioning and staining. Electron micrographs were generated using a JEOL 1200EX microscope and analysed with ImageJ software[29]. Inner and outer diameters were analysed as per ref. 30.

Assessment of telomerase activity. Telomeric repeats amplification protocol (TRAP) was combined with real-time detection of amplification products to determine telomerase activity with a Quantitative Telomerase Detection kit (US Biomax) following the manufacturer's recommendations. Total protein extract (0.5 μg) was used in each reaction. End products were resolved by PAGE in a 12.5% non-denaturing gel, stained with Sybr Green Nucleic Acid gel stain (Invitrogen) and visualized with a Bio-Rad Molecular Imager ChemiDoc System.

Cell culture and cytogenetic analysis. Ear skin fibroblasts were isolated as described previously[31]. Proliferation assays were carried in triplicate on 6-well plates. Cells were grown in RPMI-10% fetal calf serum-50 μM β-mercaptoethanol with the addition of 100 nM 4-OHT or vehicle (ethanol). Cells were counted and replated at a density of 1,000 cells per well every 4 days. Splenocytes were isolated by generating single-cell suspensions from whole spleen, stimulated for 48 h with 2.5 μg ml^{-1} concavalin A and 20 μg ml^{-1} LPS (Sigma) and treated with KaryoMAX Colcemid solution (Invitrogen) for 2 h before collection. Telomere fluorescence *in situ* hybridization (FISH) was performed on metaphase nuclei as described previously[28]. At least 15 metaphases from harvested cell cultures were analysed for telomere integrity by telomere-specific peptide nucleic acid (PNA)-FISH. Telomere signal was normalized using a Pacific Blue centromeric PNA probe. For telomere-tissue-FISH, frozen tissue sections (8-μm thick) were fixed in 2% PFA for 15 min and permeabilized in 0.5% Triton X-100 for 10 min. Two PNA probes, telomere-specific FITC-00-T2AG3 and Pacific Blue-centromere-specific, were hybridized after 4 min denaturing at 83 °C under the following conditions: 70% formamide, 0.06× SSC, 0.2% BSA, 0.5 ng μl^{-1} tRNA, 0.5 ng μl^{-1} PNA probe; overnight at 25 °C. To achieve uniform hybridization we used MAUI Mixer (BioMicro) with 40 μl chamber. Nuclei were stained with TOTO3 (Invitrogen) far red stain. Telomere signal was normalized using the centromeric PNA probe. For neurosphere assays, subventricular region of brain from 3- to 6-week-old mice was dissected, dispersed into a single-cell suspension and plated in neurobasal media (StemCellTechnologies) supplemented with EGF and bFGF (20 ng ml^{-1} each) for 4 days, in the presence of 100 nM 4-OHT or vehicle (ethanol). Primary neurospheres were dissociated and seeded at 2 cells per μl density in multi-well plates. After 7–10 days, cultures were monitored for the formation of neurospheres. Alternatively, single cells were sorted into individual wells on 384-well plates at a density of 10 cells per well on Dako MoFlo high-speed cell sorter and grown for 3 weeks. Neurospheres were transferred to culture wells coated with poly-L-ornithine (15 μg ml^{-1}) and fibronectin (1 μg ml^{-1}) and differentiated in 1% FBS in neurobasal media to measure their multipotentiality. Quantification of neurosphere numbers and diameters were performed by bright-field microscopy coupled with an in-house semi-automated segmentation algorithm generated with MATLAB software (The Mathworks). For multipotentiality assays, cells were fixed, stained with GFAP and TUJ1 antibodies and quantified as described previously[32].

Laser scanning analysis for the quantification of IHC and FISH. Laser scanning cytometry quantification was performed with an iCys Research Imaging Cytometer (Compucyte) as described earlier[32,33,34] with a few modifications. Counts of Ki67$^+$, Dcx$^+$ or Sox2$^+$ cells (DAB positive) were carried out within the subventricular zones that were predefined by a certified pathologist with the haematoxylin and eosin-stained brain architecture. The target number for each sample was approximately 500 cells counted. Olig2$^+$ cell within the corpus callosum were counted in a similar manner.

RT–PCR, Southern blotting and western blotting. DNase-treated total RNA extracted from fresh liver samples with the RNeasy kit (Qiagen) was used to prepare oligo-dT complementary DNA with Superscript III (Invitrogen). RT–PCR primers are described in Supplementary Table 3. Southern and western blots were performed following standard techniques. For western blots, 40 μg protein were loaded per lane. Antibodies used include phospho-p53 (Ser15, Cell Signaling Technologies), p21 (Santa Cruz Biotechnology), Actin (Biolegend) and horseradish peroxidase-conjugated secondary antibodies (Pierce/ThermoScientifics).

Innate olfactory avoidance test. Animals were kept in a 12-h light/dark cycle and tested in the second half of the light cycle. Male mice (30–35-week-old, $n = 4$ per experimental condition) were fasted for 20 h before testing. To avoid confounding of data owing to learning, mice were used only once. To habituate to the experimental environment, mice were placed individually in a cage that was identical to the test cage (259 × 476 × 209 mm) for 20 min before the onset of testing. Following acclimation, mice were placed in an identical test chamber where their responses were recorded on a video camera (30 frames per second, 640 × 480 pixels) mounted above the test chamber. Confounding environmental/spatial cue effects were ruled out by monitoring total time spent in different zones of the chamber in the absence of odorants. For innate avoidance tests, circular filter paper (2.5 cm diameter) was scented with 40 μl of either water or progressively higher concentrations of 2-MB (1.74×10^{-6} M to 1.74×10^{-4} M) and mouse behaviour was recorded for 3 min. Videos were transferred to computer for subsequent analysis using NoldusEthovision v3.1 behavioural analysis software[35]. First the cage was divided into thirds and then time spent in each third for the duration of the recording was determined using the software. Total distance travelled, frequency of entry into the third containing the filter paper treated with the odorants (zone 3) as well as time spent investigating in this zone (to determine innate avoidance behaviour) was collated for each treatment and genotype.

Statistical analysis. All the data were analysed by one way ANOVA with Bonferroni's post test (significantly different at $P < 0.05$). Survival curves were analysed with Mantel–Cox test.

28. Maser, R. S. *et al.* DNA-dependent protein kinase catalytic subunit is not required for dysfunctional telomere fusion and checkpoint response in the telomerase-deficient mouse. *Mol. Cell. Biol.* **27,** 2253–2265 (2007).
29. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image Processing with ImageJ. *Biophotonics Int.* **11,** 36–42 (2001).
30. Potzner, M. R. *et al.* Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system. *Mol. Cell. Biol.* **27,** 5316–5326 (2007).
31. Shao, C. *et al.* Mitotic recombination produces the majority of recessive fibroblast variants in heterozygous mice. *Proc. Natl Acad. Sci. USA* **96,** 9230–9235 (1999).
32. Paik, J. H. *et al.* FoxOs cooperatively regulate diverse pathways governing neural stem cell homeostasis. *Cell Stem Cell* **5,** 540–553 (2009).
33. Mahoney, J. E. *et al.* Quantification of telomere length by FISH and laser scanning cytometry. *Proc. SPIE* **6859,** 1–9 (2008).
34. Gorczyca, W. *et al.* Analysis of human tumors by laser scanning cytometry. *Methods Cell Biol.* **64,** 421–443 (2001).
35. Spink, A. J. *et al.* The EthoVision video tracking system–a tool for behavioral phenotyping of transgenic mice. *Physiol. Behav.* **73,** 731–744 (2001).

LETTER

The assembly of a GTPase–kinase signalling complex by a bacterial catalytic scaffold

Andrey S. Selyunin[1], Sarah E. Sutton[1], Bethany A. Weigele[1], L. Evan Reddick[1], Robert C. Orchard[1], Stefan M. Bresson[1], Diana R. Tomchick[2] & Neal M. Alto[1]

The fidelity and specificity of information flow within a cell is controlled by scaffolding proteins that assemble and link enzymes into signalling circuits[1,2]. These circuits can be inhibited by bacterial effector proteins that post-translationally modify individual pathway components[3–6]. However, there is emerging evidence that pathogens directly organize higher-order signalling networks through enzyme scaffolding[7,8], and the identity of the effectors and their mechanisms of action are poorly understood. Here we identify the enterohaemorrhagic *Escherichia coli* O157:H7 type III effector EspG as a regulator of endomembrane trafficking using a functional screen, and report ADP-ribosylation factor (ARF) GTPases and p21-activated kinases (PAKs) as its relevant host substrates. The 2.5 Å crystal structure of EspG in complex with ARF6 shows how EspG blocks GTPase-activating-protein-assisted GTP hydrolysis, revealing a potent mechanism of GTPase signalling inhibition at organelle membranes. In addition, the 2.8 Å crystal structure of EspG in complex with the autoinhibitory Iα3-helix of PAK2 defines a previously unknown catalytic site in EspG and provides an allosteric mechanism of kinase activation by a bacterial effector. Unexpectedly, ARF and PAKs are organized on adjacent surfaces of EspG, indicating its role as a 'catalytic scaffold' that effectively reprograms cellular events through the functional assembly of GTPase-kinase signalling complex.

To identify new signalling pathways targeted by bacterial pathogens, we used a human growth hormone (hGH) secretion assay[9] to measure the ability of type III and type IV effector proteins to regulate vesicle trafficking through the general secretory pathway (Fig. 1a, b). Consecutively, each bacterial effector was tagged with enhanced green fluorescent protein (eGFP) and assessed for localization at host organelles (Fig. 1b). We noted that several type III effectors encoded by the extracellular pathogen enterohaemorrhagic *E. coli* (EHEC) O157:H7 inhibited host trafficking events, whereas effectors secreted by
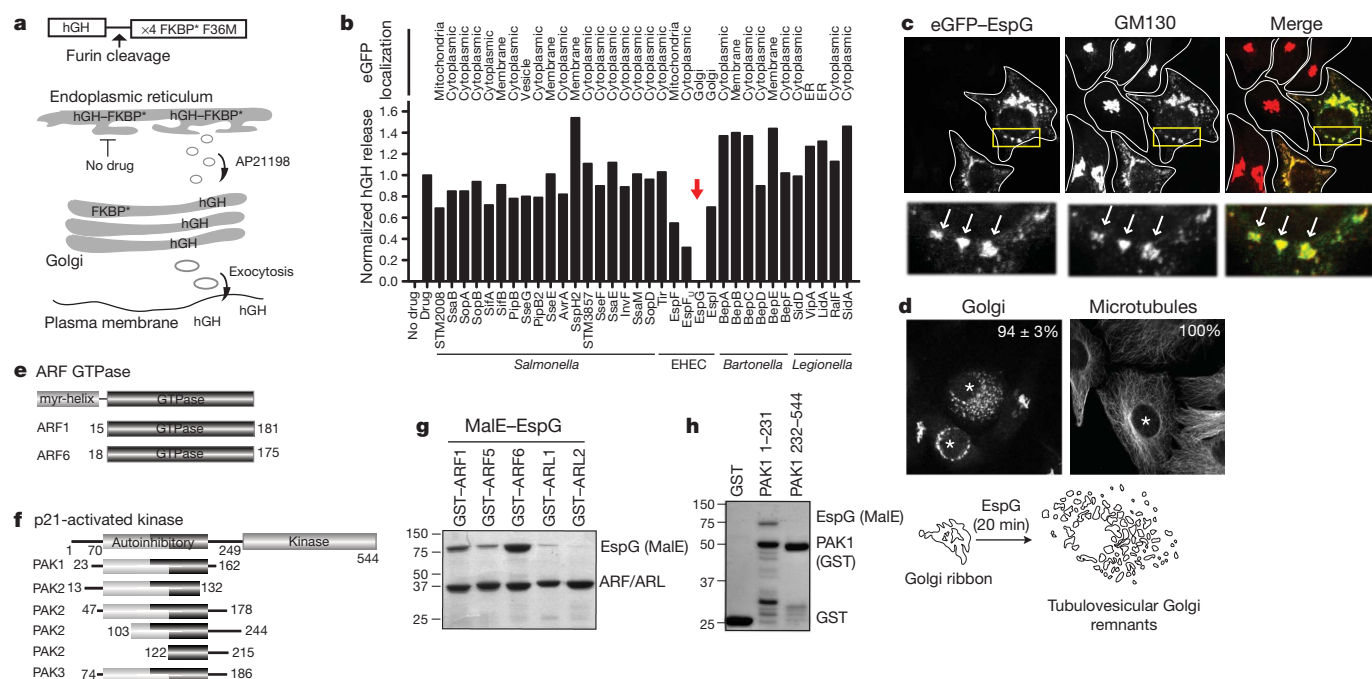

Figure 1 | EspG inhibits endomembrane trafficking and disrupts Golgi architecture. a, hGH trafficking assay showing how the hGH–FKBP* (Phe 36 Met mutant) aggregates in the endoplasmic reticulum until drug application (AP21998), whereby hGH enters the general secretory pathway and is secreted into the culture medium. **b,** hGH release assay showing the effects of type III and type IV effector proteins on trafficking through the general secretory pathway (Methods). hGH was quantified by enzyme-linked immunosorbent assay and normalized to GFP control (Drug) experiments. The subcellular localization of eGFP-tagged effectors is indicated. ER, endoplasmic reticulum. **c,** Co-localization of eGFP–EspG (green) with *cis*-Golgi matrix protein GM130 (red). The Golgi in untransfected cells appears as tightly associated cisternae. **d,** Golgi and microtubule phenotypes induced by EspG protein microinjection (asterisk). The percentage of microinjected cells exhibiting each phenotype is indicated (*n* = 3, from >40 cells per experiment). **e, f,** ARF GTPase (**e**) and PAK isoforms (**f**) that interact with EspG by yeast two-hybrid. **g, h,** Glutathione pull-down of GST–ARF isoforms (**g**) and GST–PAK1 fragments (**h**) with recombinant MalE-tagged EspG.

[1]Department of Microbiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-8816, USA. [2]Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-8816, USA.

Salmonella Typhimurium, *Legionella pneumophila* and *Bartonella henslae* showed little inhibitory functions, consistent with their intracellular life cycles (Fig. 1b). In particular, the EHEC type III effector EspG blocked exocytosis of hGH through an unknown molecular mechanism (Fig. 1b). eGFP-tagged EspG localized to the *cis*-Golgi apparatus, where it induced severe fragmentation of the organelle (Fig. 1c, d and Supplementary Fig. 1a, b). The Golgi disruption phenotype was observed when 10 nM recombinant EspG protein was microinjected into cells to mimic the protein concentration delivered by *E. coli* through the type III secretion apparatus[10] (Fig. 1d). In addition, EspG disrupted the recycling endosome compartment in both transfection (Supplementary Fig. 1c, d) and microinjection experiments (data not shown). Previous genetic studies have implicated EspG[11,12] and related *Shigella* family members[13] in microtubule depolymerization. However, microtubules were intact in EspG microinjected cells (Fig. 1d), consistent with previous reports showing that these effectors do not disrupt cytoskeletal architectures[14,15]. Thus, EspG represents a new class of bacterial signalling effector that functionally regulates cargo trafficking from membrane organelles.

We used a yeast two-hybrid screen to identify host enzymes targeted by EspG. The screen resulted in 26 positive interactions with multiple overlapping complementary DNA clones expressing two ARF GTPase family isoforms (ARF1 and ARF6) and three p21-activated kinase family members (PAK1, PAK2 and PAK3) (Fig. 1e, f). ARF GTPases function within a broad range of organelle systems, where they organize vesicle transport machinery, phospholipids and signalling molecules at membrane microdomains[16,17], whereas the PAK family of serine/threonine kinases transduce Cdc42 and Rac1 GTPase signals that establish intracellular polarity[18]. Direct interactions between EspG and the GTPase domain of ARF family members (Fig. 1g) and the autoinhibitory domain (AID) of PAK kinases (Fig. 1h) were shown by

in vitro binding studies using purified recombinant proteins. These findings establish two EspG substrates that are consistent with its regulatory function in host protein trafficking identified here and in bacterial infection studies conducted *in vivo*[19,20].

Next we crystallized EspG (residues 42–398) in complex with the GTPase domain of ARF6 (residues 13–175) and solved the structure to a resolution of 2.5 Å (Supplementary Table 1). EspG buries 602 Å^2 of ARF6 surface area and the complex interface is mediated by a collaboration of EspG loops (loops connecting β5 and β6, β8 and α6, and β12 and β13) that specifically engage the switch I loop of ARF6 and several residues lining the guanine-nucleotide-binding pocket (Fig. 2a and Supplementary Fig. 2a, b). The conformational state and amino-acid sequence of switch I are highly conserved between ARF family members, indicating that the EspG–ARF6 structure illustrates the nature of EspG's interaction with several ARF isoforms (Supplementary Fig. 3a, b). The importance of conserved switch I residues for binding EspG were confirmed by mutational analysis on ARF6 (Supplementary Fig. 3c).

ARF6 is GTP bound in the crystal and adopts an active-state conformation nearly identical to that of ARF6$_{GTPγS}$ (ref. 21; Supplementary Fig. 4a). Further structural analyses revealed that switch I is inaccessible to EspG when ARF6 adopts the GDP-bound conformation (Supplementary Fig. 4a). EspG selectively bound the GTP-loaded forms of ARF1 and ARF6 but did not recognize GDP–ARF complexes (Fig. 2b). Moreover, EspG interacted with ARF6$_{GTP}$ in its full-length myristoylated form, which was isolated from membrane fractions (Supplementary Fig. 4b). Thus, EspG preferentially targets the active ARF$_{GTP}$ signalling molecule.

COPI coat, vesicle complex adaptors and signalling enzymes primarily associate with switch 2 and the β2/3 interswitch of ARF$_{GTP}$ (refs 22–26; Supplementary Fig. 5a). Given the frequent occurrence of this
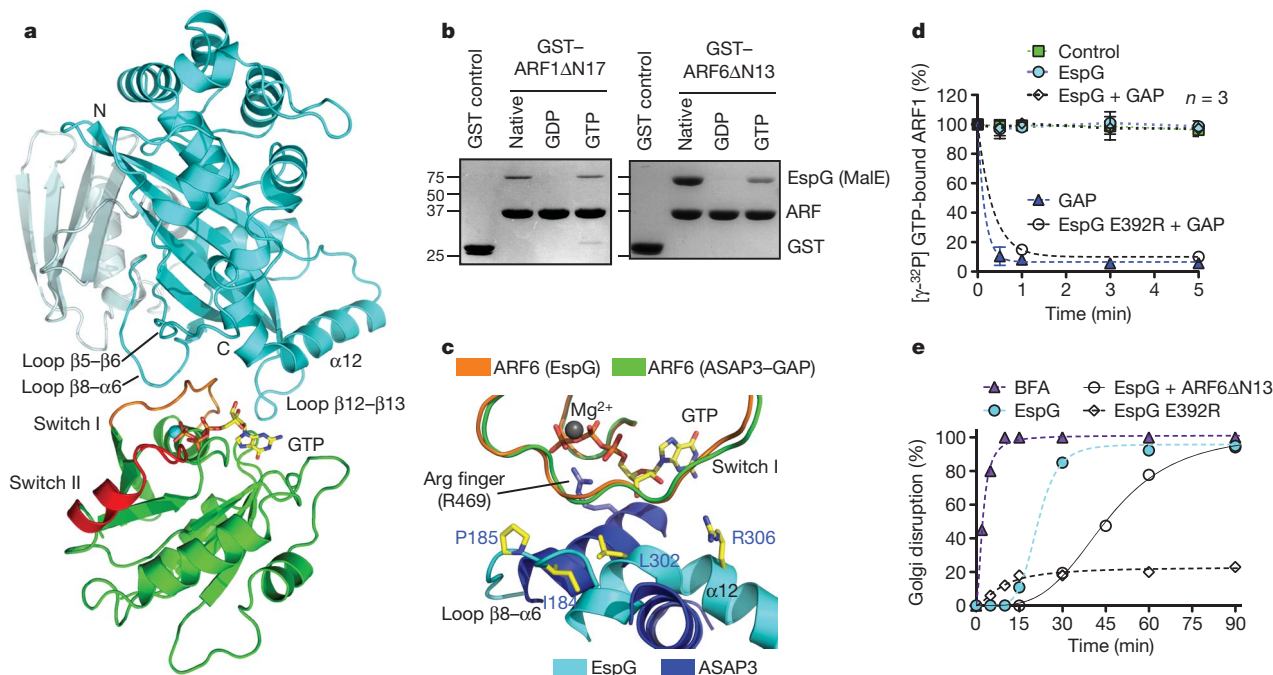

Figure 2 | The structure of EspG in complex with GTP-bound ARF6. a, The overall structure of EspG–ARF6$_{GTP}$ complex. EspG is shown in cyan and ARF6 in green. Switch I and switch II on ARF6 are coloured orange and red, respectively. **b**, EspG selectively binds the GTP-loaded ARF1 and ARF6 (GST tagged) in glutathione pull-down assays. The native lane represents ARF GTPases purified from bacteria without removing or loading specific nucleotides. **c**, Structural overlay of EspG–ARF6$_{GTP}$ and ASAP3(GAP)–ARF6$_{GDP-AlFX}$ (Protein Data Bank ID, 3LVQ) showing how EspG sterically hinders ARF binding to ASAP3–GAP. The catalytic Arg finger of ASAP3 is labelled. **d**, GTP hydrolysis assay showing that EspG inhibits GAP-assisted

GTP hydrolysis on ARF1. The rate of γ^{32}P[GTP] hydrolysis was measured as the percentage of γ^{32}P[GTP] remaining on ARF1 over time. Intrinsic ARF1 GTPase activity (control, green), GAP-stimulated activity (GAP, blue triangle), and EspG inhibition of GAP activity (EspG + GAP, open diamond) or mutant EspG Glu 392 Arg (open circle) are shown. **e**, Time course of the Golgi disruption phenotype presented as the percentage of microinjected cells with altered Golgi morphology as shown in Fig. 1d. At least 45 microinjected cells were scored in each trial for a Golgi disruption phenotype, and the data are representative of three experimental trials. BFA, brefeldin A.

binding mode, we were surprised to find that EspG is rotated away from these common binding elements and is positioned directly over the guanine-nucleotide-binding pocket (Fig. 2a, c). Surprisingly, however, EspG does not function as a guanine nucleotide exchange factor (Supplementary Fig. 5b, c) or a GTPase-activating protein (GAP) (Fig. 2d, cyan circles). Rather, EspG is appropriately positioned to hinder binding of ARF–GAP and its catalytic access to the γ-phosphate of GTP (Fig. 2c). EspG completely abolished the GAP-stimulated GTPase hydrolysis on lipid-anchored ARF1 (Fig. 2d, diamonds), in comparison with the fast ARF–GAP reaction (Fig. 2d, blue triangles). The inhibition of GAP by EspG relied on a direct interaction between EspG and ARF because the binding-deficient mutant EspG Glu 392 Arg (characterized in Supplementary Fig. 6) had no affect on GAP-stimulated hydrolysis (Fig. 2d, open circles).

GTP hydrolysis and exchange on ARF is required for proper membrane transport functions, suggesting that EspG inhibits Golgi trafficking by blocking its guanine nucleotide cycle[17,26]. Several lines of evidence support this idea. First, EspG disrupted the Golgi complex with rapid inhibitory kinetics (Fig. 2e) and a phenotype similar to the fungal toxin brefeldin A (Supplementary Fig. 7a), a potent ARF1 GTPase inhibitor that also interferes with the guanine nucleotide cycle[27]. Second, microinjection of dominant-negative ARF protein (ARFΔN13) caused a significant delay in Golgi disruption induced by EspG (Fig. 2e, open circles). Third, EspG Glu 392 Arg, a mutant that does not interact with host substrates (Supplementary Fig. 6), had no affect on Golgi morphology or trafficking function (Fig. 2e and

Supplementary Fig. 7b, c). Finally, EspG co-localized with the ARF1 effector β-COP (ref. 26) on Golgi membranes (Supplementary Fig. 1b). These combined structure and cellular studies provide a mechanism for bacterial regulation of membrane trafficking: EspG prevents vesicle transport by directly inhibiting ARF guanine nucleotide turnover on host membranes.

Having established the mechanism of ARF GTPase regulation, we next explored a second possible function of EspG: regulation of PAK family kinases. The EspG-binding site on PAK2 was defined to residues 121–136, a highly conserved sequence that encodes the Iα3-helix within the kinase AID (Supplementary Fig. 8). We crystallized EspG in complex with the PAK2 Iα3-helix fragment and solved the structure to a resolution of 2.8 Å (Supplementary Table 1). EspG recognized the initial turn of the Iα3-helix whereas the remainder of the peptide adopted an extended strand conformation that lies orthogonal to the EspG six-stranded β-sheet (Fig. 3a, b). EspG buries 684 Å^2 of the PAK2 surface area and the binding is primarily supported by a large hydrophobic interface and hydrogen bonding by residues Asn 212 and Asn 323 of EspG (Fig. 3b). This structural interface was confirmed by a series of in vitro binding studies and kinase assays using PAK2 and EspG mutant proteins (Supplementary Fig. 9).

To determine how EspG may regulate the kinase, we compared the peptide structure from EspG–PAK2 with the structure of Iα3-helix in the autoinhibited PAK1 homodimer[28] (Fig. 3c). In autoinhibited PAKs, the Iα3-helix is sandwiched between the kinase domain and the AID, where it has three autoinhibitory functions: (i) it folds onto
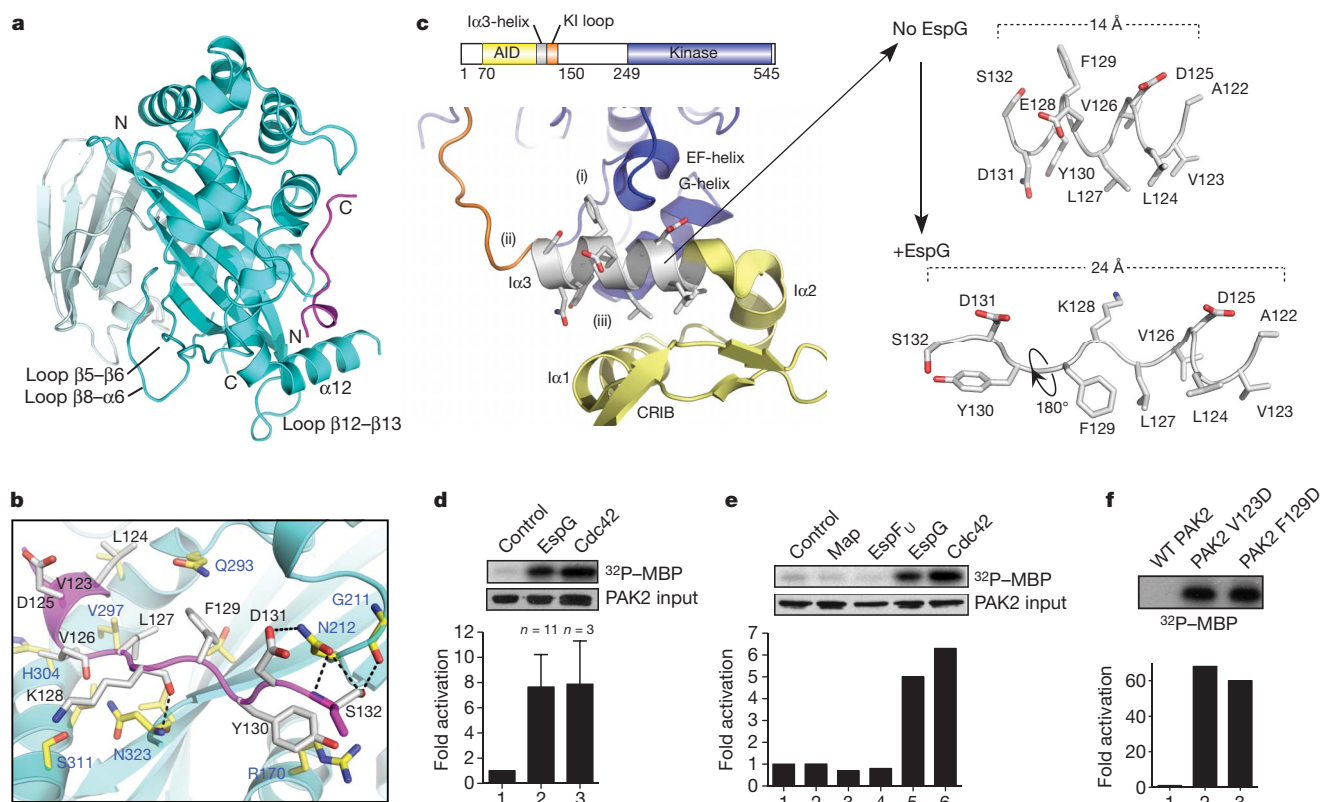

Figure 3 | The structure of EspG in complex with PAK2 Iα3 peptide. a, The overall EspG–PAK2^{Iα3} complex with EspG oriented and coloured as in Fig. 2a. The PAK2 Iα3 peptide (residues 123–134) are shown in magenta. **b**, Detailed interactions between EspG and PAK2^{Iα3}. Key binding residues from EspG (blue labels) and PAK2 (black labels) are shown. **c**, Close-up view of autoinhibited PAK1 homodimer (Protein Data Bank ID, 1F3M) focused on chain B (kinase domain, blue) and chain D (autoinhibitory domain, yellow). The Iα3-helix inhibitory functions are labelled (i)–(iii) corresponding with those outlined in the results section. The Iα3-helix extracted from the PAK1 structure (numbering corresponds to PAK2 for ease of comparison) is shown at the

upper right. The corresponding PAK2 Iα3-helix extracted from the EspG structure (lower right) is oriented by the amino-terminal helical residues 123–127. KI loop, kinase inhibitory loop. CRIB, Cdc42/Rac1 interacting binding. **d, e**, PAK2 kinase assays comparing 2 μM EspG with equimolar GTPγS-loaded Cdc42 (**d**) and the indicated EHEC type III effectors (**e**). Phosphorylation of myelin basic protein (MBP) substrate, input levels of PAK2 and quantification of each experiment are shown. **f**, PAK2 kinase assays comparing autoinhibited wild-type (WT) PAK2 with PAK2 mutants Val 123 Asp and Phe 129 Asp. Data are presented as in **d**.

the EF- and G-helices to block substrate binding; (ii) it positions the 'kinase inhibitory' loop across the enzyme catalytic cleft; and (iii) it stabilizes the AID three-helix bundle maintaining PAK homodimerization (Fig. 3c). Hence, EspG binding to the Iα3-helix suggested a mechanism for PAK activation. EspG induced a (7.6 ± 2.5)-fold ($n = 11$) increase in PAK2 activity, a profile that is comparable to PAK stimulation by GTPγS-loaded Cdc42 ((7.8 ± 3.4)-fold, $n = 3$) (Fig. 3d). Notably, EHEC type III effectors Map and EspF$_U$ showed no PAK stimulatory activity, demonstrating the signalling specificity of EspG (Fig. 3e).

To gain further insight into the mechanism of kinase activation, we first examined the details of the EspG–PAK2 interface. EspG residue Asn 212 probably initiates the kinase reaction because this residue engages the surface-accessible residues Asp 131 and Ser 132 in the autoinhibited PAK homodimer (Fig. 3b and Supplementary Fig. 10). On initial recognition, EspG displaces the Iα3-helix by reorganizing its secondary structure and by displacing side chains that normally contact the autoinhibitory interface between the kinase domain and the AID (Fig. 3c). These data suggest a novel allosteric mechanism for PAK activation. To further confirm that PAK is stimulated by local perturbations in the environment surrounding Iα3, we mutated the hydrophobic PAK2 residues Val 123 and Phe 129 that stabilize the AID and the kinase domain, respectively. Both Val 123 Asp and Phe 129 Asp resulted in a constitutively active kinase with more than 60-fold enhancement of substrate phosphorylation (Fig. 3f). We note that the mechanism of EspG binding to PAK is structurally distinct from that of Cdc42 binding[28], indicating that the catalytic machinery of EspG is a unique bacterial invention.

The two EspG structures reported here are nearly identical, with a root mean squared deviation of 0.612 Å over 349 Cα atoms. As shown in Fig. 4a, ARF6 and PAK2 occupy distinct, non-overlapping binding sites on adjacent surfaces of EspG. Consistent with this view, EspG nucleates a trimeric complex between the kinase and GTPase in solution (Supplementary Fig. 11). This complex could also be reconstituted on Golgi mimetic liposomes (Fig. 4b). ARF1$_{GTP}$ recruited EspG to the artificial membrane surface (Fig. 4b, lane 4), which in turn localized PAK2 to these sites (Fig. 4b, lane 6). Notably, PAK2 localization was strictly dependent on formation of the EspG–ARF1$_{GTP}$ complex (Fig. 4b, lanes 6–8) and ARF1 tethering to the membrane (Fig. 4b, lanes 9 and 10). As predicted by these findings, EspG co-localized with ARF1 at the Golgi (Fig. 4c). We further speculated that PAK would also be recruited to these sites. To test this hypothesis, an in vivo 'activity' probe was engineered by fusing the PAK2 Iα3-helix sequence (residues 121–136) to the carboxy terminus of the mCherry fluorophore. The PAK2 probe recognized cellular EspG and was targeted to the Golgi complex in 78 ± 5% of EspG-transfected cells (Fig. 4c). By comparison, mutant EspG Asn 212 Ala that lacked all kinase stimulatory activity (Supplementary Fig. 9) localized to the Golgi complex but did not recruit PAK2 to these sites (Fig. 4c). Together, our studies support the function of EspG as a catalytic scaffold that links GTPase inhibition with kinase signal transduction pathways at membrane organelles (Fig. 4d).

EspG belongs to large family of type III effectors secreted by diverse bacterial pathogens. Our studies show that EspG has structural homology with VirA (refs 14, 15) from *Shigella flexneri*, suggesting that it too may function as an enzyme scaffold (root mean squared deviation, 3.1 Å; Z-score, 5.9) (Supplementary Fig. 12). However, a detailed structural comparison indicates that VirA is unlikely to target the same signalling pathways as EspG during *Shigella* pathogenesis (Supplementary Fig. 12). We provide mechanistic insights and structural evidence that EspG harbours two unique pathogenic activities, ARF GTPase inhibition and PAK stimulation. Moreover, EspG targets PAK to specific membrane surfaces through its association with ARFs. From a strategic point of view, the assembly of an artificial enzyme complex enables bacteria to precisely control signalling events with little competition between endogenous regulatory systems. Thus, it is intriguing to speculate that EspG organizes a higher-order signalling network to effectively subvert key cellular processes including cell polarity, adhesion, receptor trafficking and protein secretion.
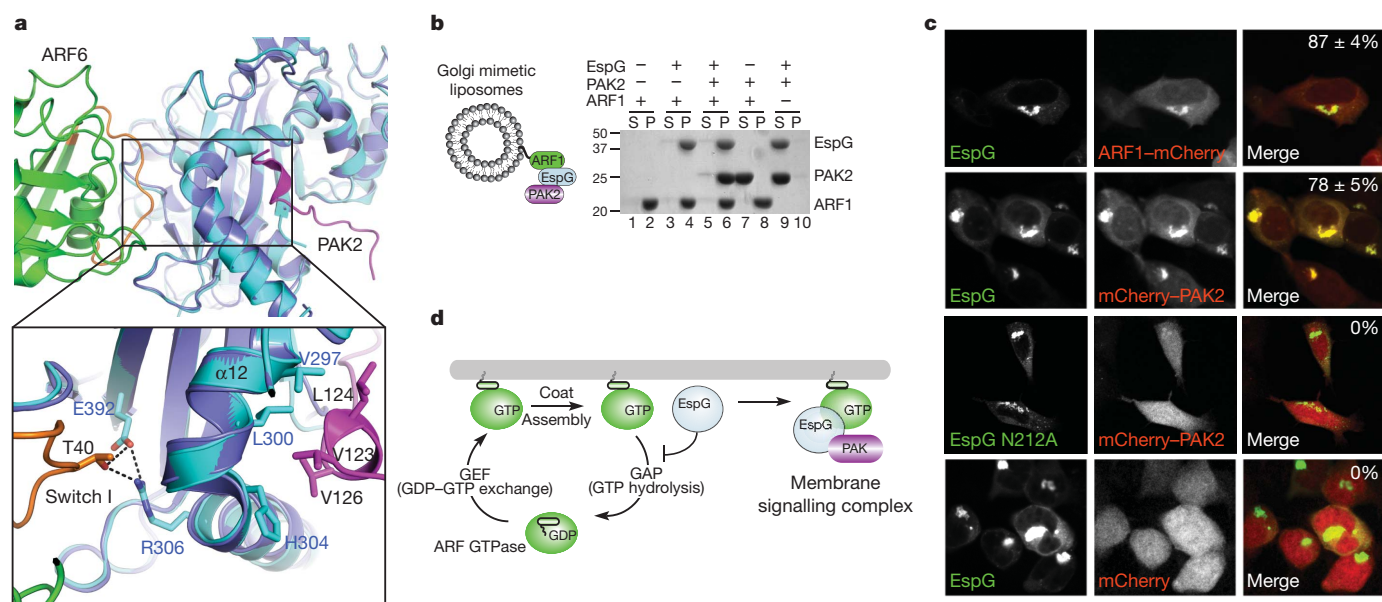

Figure 4 | EspG functions as a catalytic scaffold at membrane organelles.
a, Structural overlay of EspG–ARF6$_{GTP}$ and EspG–PAK2^{Iα3} highlighting the close association between ARF and PAK on the surface of EspG. Colours are as in Figs 2a and 3a except that EspG from the PAK2 structure is coloured purple. **b**, Golgi-mimetic-liposome-binding assays showing that EspG nucleates a trimeric complex between ARF1 and PAK2 on membrane surfaces. After centrifugation, proteins remaining in the supernatant (S) or those associated with liposomes in the pellet (P) are indicated. **c**, HEK239A cells co-transfected with the indicated constructs showing that eGFP–EspG co-localizes with ARF1–mCherry and recruits a PAK activity probe (mCherry–PAK2$^{121–136}$) to Golgi membranes. The percentage of cells exhibiting co-localized EspG with mCherry-tagged proteins ($n = 3$) is shown in the upper right of the merged micrographs. **d**, Model of the dual function of EspG as an inhibitor of membrane trafficking and as a catalytic scaffold that assembles a GTPase–kinase signalling complex at cellular membranes. GEF, guanine nucleotide exchange factor.

METHODS SUMMARY

Recombinant protein preparation and cloning were done using standard methods. In immunofluorescence experiments, we transfected cells with Fugene 200 or microinjected them with 10 nM protein, where indicated. We performed kinase assays with rabbit pEGFP–PAK2, which was immunoprecipitated from 293T cell lysates and incubated with the protein of interest in the presence of MBP, 10 μM ATP and 5 μCi γ^{32}P[ATP]. Reactions were stopped by the addition of SDS buffer, separated by SDS–polyacrylamide gel electrophoresis and kinase activity was measured as ^{32}P counts per minute. EspG–ARF6 and EspG–PAK2–Iα3 were purified by ion exchange and gel filtration chromatography and crystallized by the hanging-drop vapour diffusion method. We collected X-ray diffraction data at the Structural Biology Center, Advanced Photon Source, Argonne National Laboratory (USA). The structure of EspG–ARF6 was phased to a resolution of 2.5 Å by the multiwavelength anomalous dispersion method using selenomethionine-labelled EspG and ARF6 proteins. The EspG–PAK2 structure was solved to a resolution of 2.8 Å by the molecular replacement method using the EspG monomer of the EspG–ARF6 structure as the initial search model. Further details can be found in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Scott, J. D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326,** 1220–1224 (2009).
2. Lim, W. A. Designing customized cell signalling circuits. *Nature Rev. Mol. Cell Biol.* **11,** 393–403 (2010).
3. Duesbery, N. S. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280,** 734–737 (1998).
4. Yarbrough, M. L. *et al.* AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science* **323,** 269–272 (2009).
5. Schmidt, G. *et al.* Gln 63 of Rho is deamidated by *Escherichia coli* cytotoxic necrotizing factor-1. *Nature* **387,** 725–729 (1997).
6. Li, H. *et al.* The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **315,** 1000–1003 (2007).
7. Alto, N. M. *et al.* The type III effector EspF coordinates membrane trafficking by the spatiotemporal activation of two eukaryotic signaling pathways. *J. Cell Biol.* **178,** 1265–1278 (2007).
8. Vingadassalom, D. *et al.* Insulin receptor tyrosine kinase substrate links the *E. coli* O157:H7 actin assembly effectors Tir and EspF$_U$ during pedestal formation. *Proc. Natl Acad. Sci. USA* **106,** 6754–6759 (2009).
9. Rivera, V. M. *et al.* Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science* **287,** 826–830 (2000).
10. Winnen, B. *et al.* Hierarchical effector protein transport by the Salmonella Typhimurium SPI-1 type III secretion system. *PLoS ONE* **3,** e2178 (2008).
11. Shaw, R. K. *et al.* Enteropathogenic *Escherichia coli* type III effectors EspG and EspG2 disrupt the microtubule network of intestinal epithelial cells. *Infect. Immun.* **73,** 4385–4390 (2005).
12. Tomson, F. L. *et al.* Enteropathogenic *Escherichia coli* EspG disrupts microtubules and in conjunction with Orf3 enhances perturbation of the tight junction barrier. *Mol. Microbiol.* **56,** 447–464 (2005).
13. Yoshida, S. *et al.* Microtubule-severing activity of *Shigella* is pivotal for intercellular spreading. *Science* **314,** 985–989 (2006).
14. Germane, K. L., Ohi, R., Goldberg, M. B. & Spiller, B. W. Structural and functional studies indicate that *Shigella* VirA is not a protease and does not directly destabilize microtubules. *Biochemistry* **47,** 10241–10243 (2008).
15. Davis, J. *et al.* Novel fold of VirA, a type III secretion system effector protein from *Shigella flexneri. Protein Sci.* **17,** 2167–2173 (2008).
16. Kahn, R. A. Toward a model for Arf GTPases as regulators of traffic at the Golgi. *FEBS Lett.* **583,** 3872–3879 (2009).
17. D'Souza-Schorey, C. & Chavrier, P. ARF proteins: roles in membrane traffic and beyond. *Nature Rev. Mol. Cell Biol.* **7,** 347–358 (2006).
18. Bokoch, G. M. Biology of the p21-activated kinases. *Annu. Rev. Biochem.* **72,** 743–781 (2003).
19. Borthakur, A. *et al.* Enteropathogenic *Escherichia coli* inhibits butyrate uptake in Caco-2 cells by altering the apical membrane MCT1 level. *Am. J. Physiol. Gastrointest. Liver Physiol.* **290,** G30–G35 (2006).
20. Guttman, J. A. *et al.* Aquaporins contribute to diarrhoea caused by attaching and effacing bacterial pathogens. *Cell. Microbiol.* **9,** 131–141 (2007).
21. Pasqualato, S., Menetrey, J., Franco, M. & Cherfils, J. The structural GDP/GTP cycle of human Arf6. *EMBO Rep.* **2,** 234–238 (2001).
22. Hanzal-Bayer, M., Renault, L., Roversi, P., Wittinghofer, A. & Hillig, R. C. The complex of Arl2-GTP and PDE delta: from structure to function. *EMBO J.* **21,** 2095–2106 (2002).
23. Isabet, T. *et al.* The structural basis of Arf effector specificity: the crystal structure of ARF6 in a complex with JIP4. *EMBO J.* **28,** 2835–2845 (2009).
24. O'Neal, C. J., Jobling, M. G., Holmes, R. K. & Hol, W. G. Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science* **309,** 1093–1096 (2005).
25. Shiba, T. *et al.* Molecular mechanism of membrane recruitment of GGA by ARF in lysosomal protein transport. *Nature Struct. Biol.* **10,** 386–393 (2003).
26. Zhao, L. *et al.* Direct and GTP-dependent interaction of ADP ribosylation factor 1 with coatomer subunit beta. *Proc. Natl Acad. Sci. USA* **94,** 4418–4423 (1997).
27. Chardin, P. & McCormick, F. Brefeldin A: the advantage of being uncompetitive. *Cell* **97,** 153–155 (1999).
28. Lei, M. *et al.* Structure of PAK1 in an autoinhibited conformation reveals a multistage activation switch. *Cell* **102,** 387–397 (2000).

METHODS

Plasmids. The *espG* gene from EHEC O157:H7 was PCR cloned in-frame into pEGFP-C2 (Clontech) and pcDNA3.1-mCherry. For bacterial expression, 38 and 41 amino-acid N-terminal deletions (39–398 and 42–398) of EspG were PCR subcloned into pGEX-4T1 (GST-tag) (Amersham), pProEX-HTb (6xHis tag) (Novagen) and pET28b-MalE (6xHis tag, MalE-tag) vectors. EspG mutants were generated with QuickChange site-directed mutagenesis (Stratagene) following manufacturer's instructions. N-terminal deletions of ARF GTPases (ARF1Δ17, ARF5Δ17 and ARF6Δ13) and ARL proteins (ARL1Δ17 and ARL2Δ16) were PCR subcloned into pProEX-HTb vectors. Human PAK1 construct was obtained from Dr Gary Bockoch (TSRI, La Jolla, California), and rabbit PAK2 and PAK3 were obtained from Dr Melanie Cobb (UTSW). PCR cloning was used to generate variable-length constructs of PAK isoforms in pGEX-4T1 vector. All constructs were verified by DNA sequencing.

Yeast two-hybrid system. The yeast expression vector pLexA encoded a gene with an NH2-terminal LexA-binding domain and residues 1–398 of EHEC EspG. Day 9.5 and 10.5 mouse embryo cDNA library (250 μg) in VP16 were screened using the yeast two-hybrid system.

Protein purification for *in vitro* assays. Recombinant proteins were produced in BL21-DE3 *E. coli* strains. Protein expression was induced with 0.4 mM IPTG for 16 h at 18 °C. Bacterial pellets were lysed in either His buffer (100 mM HEPES, pH 7.5, 300 mM NaCl) or GST buffer (TBS; 50 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT) supplemented with protease cocktail (Roche). Proteins were purified with nickel agarose (Qiagen) or glutathione Sepharose (Amersham Biosciences) following manufacturer's instructions. Eluted proteins were buffer exchanged into TBS using concentration centrifugal columns (Millipore), glycerol was added to 15% and the proteins were then stored at −80 °C.

In vitro GST pull-downs. Protein interactions were examined through GST pull-down assays. Unless otherwise stated, 15 μg of recombinant GST proteins immobilized to glutathione Sepharose were incubated with 20 μg of 6xHis- and/or MalE-tagged proteins for 1 h at 4 °C. Samples were washed three times in TBS supplemented with 0.5% Triton X-100. Proteins were eluted from beads with Laemelli sample buffer and were separated by SDS–polyacrylamide gel electrophoresis and stained with Coomassie blue. For nucleotide loading, ARF1Δ17 and ARF6Δ13 were incubated in nucleotide loading buffer (40 mM HEPES, 150 mM NaCl, 2 mM EDTA, 10% glycerol) with 10 μM of either GDP or GTP for 30 min at 37 °C, and then MgCl$_2$ was added to 10 mM and the reaction was transferred to ice after 15 min at room temperature (25 °C).

Kinase assays. To obtain full-length PAK2 kinase, 10-cm dishes with 293A cells were transfected with 5 μg rabbit PAK2 cDNA in pEGFP-C2 vector and expressed for 48 h post-transfection. Cells were broken in lysis buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% Triton X-100). PAK2 kinase was purified by immunoprecipitation with 1:1,000 polyclonal anti-GFP antibody (Clontech) and 25 μl protein A/G slurry for 1 h at 4 °C. Beads were washed twice with lysis buffer and twice with kinase buffer (40 mM HEPES, pH 7.5, 10 mM MgCl$_2$). MBP (5 μg) and 2 μM activating proteins (that is, EspG or Cdc42) were added to the beads for a total volume of 30 μl. The reaction was equilibrated on ice for 30 min. The kinase activity was initiated by an addition of 10 mM ATP and 5 uCi ATPγP^{32}. After 5 min at room temperature, the reaction was stopped with 30 μl ×2 Laemelli sample buffer. Contents were separated by SDS–polyacrylamide gel electrophoresis, transferred to nitrocellulose membrane and either analysed by western blot (1:5,000 monoclonal anti-GFP) or exposed by autoradiography. Bands were cut out and the radioactivity signal measured on a scintillation counter.

Cell microinjection, transfections and immunofluorescence microscopy. Normal rat kidney cells were microinjected with EspG proteins using a semi-automatic InjectMan NI2 micromanipulator (Eppendorf). A needle concentration of 10 nM was calculated to inject between 5,000 and 20,000 copies because we microinjected ~5% cell volume, giving a final estimated cellular concentration of 50 pM in a cell volume of 5,000 μm^3. HeLa and HEK293A cells were transfected using calcium phosphate. At 16–18 h post-transfection, cells were fixed with 3.7% formaldehyde and stained with antibodies for immunofluorescence. In co-transfection experiments, equal amounts of DNA were used for each sample. Brefeldin A treatment was performed by adding 5 μg ml^{-1} of brefeldin A to the medium before fixation with formaldehyde. As a negative control, ethanol was added to the medium. All immunofluorescence images were acquired with a Zeiss LSM 5 Pascal confocal microscope. Golgi, endosomes and microtubules were detected using anti-GM130 (transduction labs), anti-EEA1 (transduction labs) and anti-α-tubulin (Sigma) antibodies, respectively.

hGH trafficking assay. hGH trafficking assays were performed as described previously[9]. Briefly, HeLa cells (50% confluence) were transfected with 1 μg of 4xFKBP-hGH (Ariad Pharmaceuticals, Inc.; http://www.ariad.com/regulationkits; source of material, David Bernstein) and either 0.5 μg eGFP–EspG or pEGFP control plasmid with Fugene6 (Roche). Sixteen hours later, the medium was

replaced with medium containing AP21998 (final concentration, 2 μM) or vehicle control. AP21998 was incubated with the cells for 2 h before the supernatant was collected. The supernatant was then diluted 100-fold and compared against a hGH standard curve (12.5–400 pg ml^{-1}) for the quantification of hGH released using an hGH enzyme-linked immunosorbent assay (Roche). For no drug controls, 100% ethanol (2 μl) was incubated with the cells for 2 h.

Liposome pull-downs and GAP assays. Liposome preparation: Lipids were purchased in powder form from Avanti Polar Lipids. Golgi mimetic liposomes were created by combining 20 mol% DOGS-NTA with DOPC, DOPE, DOPA, DOPS, PI, PI$_4$P and PI$_{4,5}$P$_2$ in the molar ratios reported previously[29]. Total lipid (5 mM) was solubilized in chloroform, dried under an anhydrous nitrogen stream and further dried in a vacuum desiccator for approximately 5 h. Dried lipids were hydrated with liposome-binding buffer (20 mM Tris-HCl, pH 7.6, 50 mM NaCl, 10 mM MgCl$_2$) and vigorously vortexed between five freeze–thaw cycles (liquid nitrogen and 80 °C to ensure appropriate phase transition and dispersion of the various lipids), after which liposomes were generated by means of ultrasonication in a bath sonicator (Laboratory Supplies Company). Liposomes were collected from the supernatant after centrifugation (2,500g for 5 min) and used in subsequent assays.

Liposome pull-down assays: Liposomes were prepared as described above. ARF1 GTP loading was carried out by incubating purified 6xHis-tagged ARF1ΔN17 in nucleotide exchange buffer (20 mM Tris-HCl, pH 7.6, 50 mM NaCl, 5 mM EDTA, 10% glycerol, 1 mM DTT) with 100 μM GTP. After incubation at 37 °C for 30 min, 10 mM MgCl$_2$ was added to stabilize ARF1(GTP). The requisite volume of ARF1(GTP), GST PAK (residues 121–136) or EspG was added to bring the protein concentration to 3 μM in a 100-μl volume. Liposomes were added for a final lipid concentration of 10 μM and reactions proceeded at room temperature for 30 min. Samples were subjected to centrifugation at 100,000g in a Beckman TLA100.3 rotor and a Beckman TL100 ultracentrifuge at 4 °C for 1 h. Supernatant and pellet were separated and analysed on a 12.5% polyacrylamide gel and visualized with Coomassie blue.

GTP hydrolysis assays: ARF1 was incubated in nucleotide exchange buffer with 250 nM γ^{32}P[GTP] (MP Biomedical) for 30 min at 37 °C, after which 10 mM MgCl was added to stabilize ARF1(γ^{32}P[GTP]). ARF1(γ^{32}P[GTP]) was incubated with 10 μM Golgi mimetic liposomes 5 min before the addition of fivefold molar excess rat ARFGAP1 (ref. 30), EspG and EspG(E392R). In the case of hydrolytic protection assays, EspG or EspG(E392R) was added 5 min before the addition of rat ARFGAP1. Aliquots (5 μl) of the 50-μl reaction were removed at times indicated, added to 5 ml ice-cold binding buffer (TBS + 10 mM MgCl$_2$) and vacuum-filtered through nitrocellulose membranes. Membranes were washed three times with ice-cold binding buffer and subjected to scintillation counting. Data analysis was carried out in GRAPHPAD PRISM 5.0b.

Crystallization and structure determination. Protein expression and purification: A stable protein fragment of EspG residues 42–398 was identified by limited proteolysis and mass spectrometry. cDNA-encoding EHEC O157:H7 EspG residues 42–398 and human ARF6 residues 14–175 were synthesized by PCR and ligated into the pPRO-EX-HTb expression vector. The resulting plasmids were then transformed into the *E. coli* strain BL21-DE3. Protein expression was induced by 1 mM IPTG overnight at 16 °C and proteins were purified on Ni-NTA agarose, concentrated to 10 mg ml^{-1} in TBS buffer with 5% glycerol, snap-frozen in liquid nitrogen and stored at −80 °C. The Se-Met variant of EspG and ARF6 was expressed in methionine auxotrophic *E. coli* strain B834-DE3 and grown in minimal medium supplemented with natural amino acids and Se-Met. Expression and purification were unchanged. EspG–ARF6 complex was formed overnight at room temperature in the presence of 1:100 TEV protease to cleave the 6xHis tag. The complex was purified by successive anion exchange (Q-HP) and gel filtration (Superdex 200 GL) chromatography and concentrated to 7 mg ml^{-1} in 25 mM Tris, pH 7.5, and 50 mM NaCl. For the PAK2 crystal trials, EspG protein was expressed and purified identically to that described. However, after the anion exchange, EspG was incubated with a fivefold molar excess of PAK2 peptide (residues 121–136). The complex was purified by gel filtration as described.

Crystallization and X-ray diffraction data collection: Crystals of EspG–ARF6 were grown using the hanging-drop vapour diffusion method from drops containing 2 μl protein (7 mg ml^{-1}) and 1 μl of reservoir solution (0.1 M sodium acetate, pH 5.0, 2% PEG4000, 5% 2,3-methylpentanediol (MPD)), and equilibrated over 500 μl of reservoir solution. Bipyramid-like crystals appeared after 1 d at 20 °C and grew to their maximal extent in 2–3 d. Crystals were relatively large in all three dimensions (0.3 × 0.6 × 0.3 mm^3). Cryo-protection was performed by transferring the crystals to a final solution of 37% MPD, 0.1 M sodium acetate, pH 5.0, and 2% PEG4000, increasing in 5% steps of MPD over the course of 10 min at 20 °C. Crystals were flash-frozen using liquid nitrogen. EspG–ARF6 crystals had the symmetry of space group $P4_32_12$ with unit-cell parameters of $a = b = 104.6$ Å and $c = 98.3$ Å, and contained one molecule each of EspG and

ARF6 per asymmetric unit. EspG–ARF6 crystals diffracted isotropically to a d_{min} of 2.50 Å when exposed to synchrotron radiation.

Crystals of EspG–PAK2 were grown using the hanging-drop vapour diffusion method from drops containing 1 μl protein (12 mg ml^{-1}) and 1 μl of reservoir solution (0.1 M Tris, pH 8.0, 0.25 M sodium chloride and 20% PEG4000) and equilibrated over 500 μl of reservoir solution. Plate-like crystals appeared after 2 d at 20 °C and grew to their maximal extent by 4–5 d. Crystals were large in two dimensions (0.2×0.5 mm^2) and relatively thin (0.1 mm). Cryo-protection was performed by transferring the crystals to a final solution of 15% ethylene glycol, 22% PEG4000, 0.1 M Tris, pH 8.0, and 0.25 M sodium chloride, increasing in 5% steps of ethylene glycol over the course of 10 min at 20 °C. Crystals were flash-frozen using liquid nitrogen. EspG–PAK crystals had the symmetry of space group $P2_12_12_1$ with unit-cell parameters of $a = 86.7$ Å, $b = 104.6$ Å and $c = 192.0$ Å, and contained four molecules of EspG–PAK per asymmetric unit. EspG–PAK crystals diffracted to a d_{min} of 2.85 Å when exposed to synchrotron radiation. Data were indexed, integrated and scaled using the HKL-3000 program package[31]. Data collection statistics are provided in Supplementary Table 1.

Phase determination and structure refinement: Phases for the EspG–ARF6 complex were obtained from a three-wavelength anomalous dispersion experiment using selenomethionyl-substituted protein with data to a d_{min} of 2.50 Å. Fifteen selenium sites were located using the program SHELXD[32]; this represented nine single-occupancy selenium sites and six half-occupancy selenium sites per EspG–ARF6 heterodimer. Phases were refined with the program MLPHARE, resulting in an overall figure of merit of 0.41 for data between 32.9 and 2.50 Å. Phases were further improved by density modification with the program DM[33], resulting in a figure of merit of 0.70. An initial model containing 97% of all EspG residues was automatically generated by alternating cycles of the programs RESOLVE[34] and BUCCANEER[35]. Inspection of the electron density map revealed density for the ARF6 molecule, but the automatic model-building programs were unable to build a complete model for this protein. Placement of a model for ARF6 in the cell was performed by means of molecular replacement in the program PHASER[36] using the GTPγS-bound ARF6 (Protein Data Bank ID, 2J5X) as a search model.

Additional residues for EspG were manually modelled in the program O[37]. Refinement was performed with the data collected at the selenium peak wavelength to a resolution of 2.50 Å using the program PHENIX[38] with a random 5% of all data set aside for an R_{free} calculation. The current model contains one EspG and one ARF6 monomer; included are residues 47–395 of EspG, residues 14–174 of ARF6, one Mg^{2+}-GTP and 138 water molecules. The R_{work} value is 22.5% and the R_{free} value is 32.4%. The higher-than-average R_{free} value is probably due to the relative dearth of lattice contacts for the ARF6 molecule, as evidenced by weak electron density for the portions of ARF6 that are distal to the EspG-binding site.

The density for the portions of ARF6 (residues 20–63 and 152–170) that are proximal to EspG and to the Mg^{2+}-GTP is strong and well connected. A Ramachandran plot generated with MOLPROBITY[39] indicated that 99.0% of all protein residues are in allowed regions.

Phases for EspG–PAK were obtained by means of molecular replacement in the program PHASER using the coordinates of EspG from the EspG–ARF6 structure as a search model. Model building and refinement were performed as described above, with the following modification: owing to the lower resolution of the data, restrained non-crystallographic symmetry was implemented during refinement. The current model contains four EspG monomers and four PAK peptides. Included are EspG residues 42–158, 163–318 and 321–397 and PAK residues 122–135, in complex A; EspG residues 43–397 and PAK residues 122–133, in complex B; EspG residues 42–158, 163–316 and 322–395 and PAK residues 123–132, in complex C; and EspG residues 42–158, 163–317 and 320–395 and PAK residues 122–134, in complex D. The R_{work} value is 20.3% and the R_{free} value is 28.6%. A Ramachandran plot generated with MOLPROBITY indicated that 99.4% of all protein residues are in allowed regions. Phasing and model refinement statistics are provided in Supplementary Table 1.

Combination models were generated by structural alignment of homologous or identical proteins from separate independent structures where applicable, using PYMOL.

29. Bremser, M. *et al.* Coupling of coat assembly and vesicle budding to packaging of putative cargo receptors. *Cell* **96,** 495–506 (1999).
30. Goldberg, J. Structural and functional analysis of the ARF1-ARFGAP complex reveals a role for coatomer in GTP hydrolysis. *Cell* **96,** 893–902 (1999).
31. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62,** 859–866 (2006).
32. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
33. Cowtan, K. & Main, P. Miscellaneous algorithms for density modification. *Acta Crystallogr. D* **54,** 487–493 (1998).
34. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11,** 49–52 (2004).
35. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62,** 1002–1011 (2006).
36. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
37. Jones, T. A. Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
38. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
39. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).

LETTER

CENP–B preserves genome integrity at replication forks paused by retrotransposon LTR

Mikel Zaratiegui[1], Matthew W. Vaughn[1]†, Danielle V. Irvine[1]†, Derek Goto[1]†, Stephen Watt[2]†, Jürg Bähler[2], Benoit Arcangioli[3] & Robert A. Martienssen[1]

Centromere-binding protein B (CENP-B) is a widely conserved DNA binding factor associated with heterochromatin and centromeric satellite repeats[1]. In fission yeast, CENP-B homologues have been shown to silence long terminal repeat (LTR) retrotransposons by recruiting histone deacetylases[2]. However, CENP-B factors also have unexplained roles in DNA replication[3,4]. Here we show that a molecular function of CENP-B is to promote replication-fork progression through the LTR. Mutants have increased genomic instability caused by replication-fork blockage that depends on the DNA binding factor switch-activating protein 1 (Sap1), which is directly recruited by the LTR. The loss of Sap1-dependent barrier activity allows the unhindered progression of the replication fork, but results in rearrangements deleterious to the retrotransposon. We conclude that retrotransposons influence replication polarity through recruitment of Sap1 and transposition near replication-fork blocks, whereas CENP-B counteracts this activity and promotes fork stability. Our results may account for the role of LTR in fragile sites, and for the association of CENP-B with pericentromeric heterochromatin and tandem satellite repeats.

In fission yeast, CENP-B proteins are encoded by three homologues, autonomously replicating sequence binding protein 1 (abp1), cenp-B homologue 1 (cbh1) and cbh2, and were previously characterized as DNA binding factors at origins of replication and centromeric repeats, respectively[3,5]. Mutants of abp1 grow slowly, whereas double mutants with cbh1 or cbh2 have severely stunted growth, abnormal mitosis and morphological defects, and triple deletion mutants are inviable[6,7]. As a result, double Δabp1Δcbh1 mutants form microcolonies on solid

media (Fig. 1a and Supplementary Table 1) and exhibit high levels of cell death (Supplementary Fig. 1). We observed the spontaneous appearance of faster growing cells in a culture of Δabp1Δcbh1 that grew at rates similar to the Δabp1 single mutant, lacked morphological defects (Fig. 1a and Supplementary Table 1) and showed lower levels of cell death (Supplementary Fig. 1). Genetic analysis revealed the presence of a single essential locus that also suppressed the lethality of the triple mutant Δabp1Δcbh1Δcbh2 (not shown). We performed whole-genome resequencing in the mutant strain[8] and isolated a missense mutation in the coding sequence of the DNA binding factor Sap1 (sap1E101D, henceforth called sap1-c; Supplementary Fig. 2) that co-segregated with suppression of slow growth in Δabp1Δcbh1 and resulted in lethality in a wild-type background. Sap1 is a protein with essential roles in chromosome stability[9]. Sap1 has been implicated in a programmed replication-fork block in the ribosomal DNA (rDNA) monomer that ensures directional replication to prevent mitotic recombination between rDNA repeats[10–12].

To test the effects of CENP-B and sap1-c mutations on genome integrity, we examined chromosomes by pulsed-field gel electrophoresis. Although single Δabp1 and Δcbh1 mutants had wild-type chromosome lengths, the double Δabp1Δcbh1 mutant had a smear of DNA fragments indicating double-strand breaks in all three chromosomes (Fig. 1b). Treatment of the Δabp1Δcbh1 sample plugs with the restriction enzyme NotI allowed migration of the chromosomes into the gel, and detection of telomeric and centromeric sequences (Supplementary Fig. 3), suggesting the presence of scattered unresolved replication or recombination intermediates that interfere with the migration of full-length chromosomes,
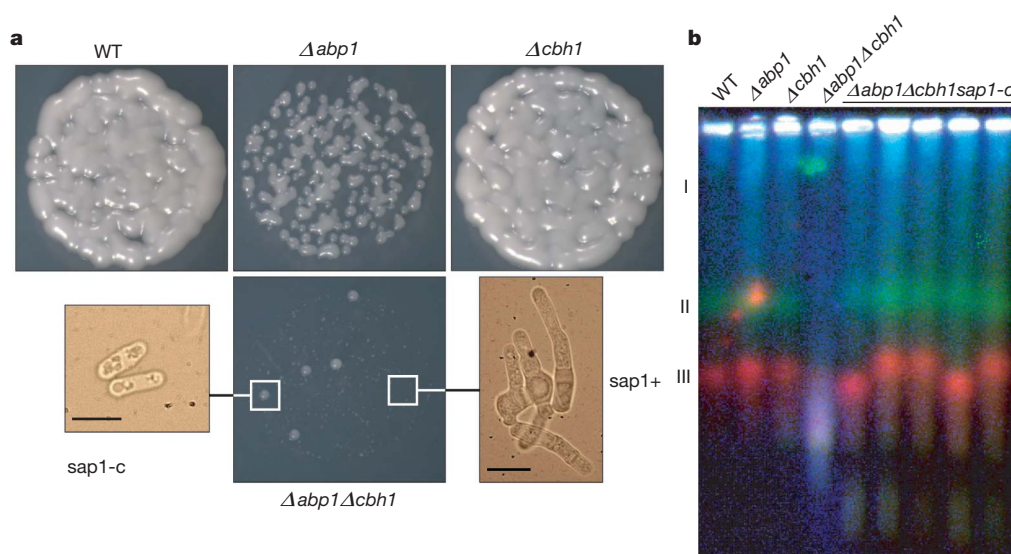

Figure 1 | DNA damage in CENP-B mutants is suppressed by sap1 mutation. a, Images of 10³ plated cells of wild type (WT), Δabp1, Δcbh1 and Δabp1Δcbh1 with Δabp1Δcbh1sap1-c colonies. Microscopy image inserts: branched phenotype in Δabp1Δcbh1 background (right) and Δabp1Δcbh1sap1-c mutant (left). Scale bar, 10 μm. b, Pulsed-field gel blot analysis of WT, CENP-B mutants (Δabp1, Δcbh1, Δabp1Δcbh1) and five CENP-B/ sap1-c mutant isolates (Δabp1Δcbh1sap1-c). The position of the three chromosomes is indicated on the left. The image is a false-coloured composite of hybridizations for all three chromosomes.

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]University College London, Department of Genetics, Evolution & Environment, and UCL Cancer Institute, Darwin Building, Gower Street, London WC1E 6BT, UK. [3]Institut Pasteur, Dynamics of the Genome Unit, Departments of Genomes and Genetics and Developmental Biology, CNRS URA2171, F-75015 Paris, France. †Present addresses: Texas Advanced Computing Center, University of Texas at Austin, 10100 Burnet Road, Austin, Texas 78758, USA (M.W.V.); Chromosome and Chromatin Research, Murdoch Childrens Research Institute, Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Flemington Road, Parkville 3052, Australia (D.V.I.); Creative Research Initiative Sousei, Hokkaido University. North-21, West-10, Kita-ku, Sapporo, 001-0021, JAPAN (D.G.); Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK (S.W.).

but not with NotI-digested DNA, into the pulsed field gel. This indicates that Abp1 and Cbh1 have roles in the maintenance of genome integrity.

Surprisingly, the *sap1-c* mutation restored genome integrity to all chromosomes, with chromosome 3 exhibiting size variability in several isolates of *Δabp1Δcbh1sap1-c* mutant (Fig. 1b). In fission yeast, chromosome 3 harbours the rDNA repeats. Temperature-sensitive alleles of *sap1* exhibit changes in the size of chromosome 3 attributable to loss of fork barrier activity and an increase in mitotic recombination at rDNA[12]. The changes in the size of chromosome 3 size in the *sap1-c* mutants are associated with altered rDNA copy number (Supplementary Fig. 4). The temperature-sensitive alleles *sap1-1* and *sap1-48* (ref. 12) suppressed slow growth in the *Δabp1Δcbh1* double mutant, mimicking *sap1-c* (Supplementary Fig. 5). Consistent with a reduction in fork barrier activity, a probe containing a canonical Sap1-binding sequence had reduced electrophoretic mobility shift in crude extracts from *sap1-c* mutants (Supplementary Fig. 6). We conclude that the suppression of the *Δcbh1Δabp1* phenotype is not specific to the *sap1-c* mutation but a result of defective function of Sap1, and therefore that the loss of genome integrity in *Δabp1Δcbh1* mutants is a consequence of Sap1 activity.

Blocked replication forks are potential sources of genome instability because they can lead to collapse of the replisome and double-strand break formation[13]. The fact that Sap1 activity leads to DNA damage in the absence of Abp1/Cbh1 suggests that the function of CENP-B is to manage Sap1-arrested replication forks. In the absence of Sap1, loss of replication-fork blockage would render Abp1/Cbh1 activity unnecessary and lead to increased genome stability in *Δabp1Δcbh1* mutants. This model predicts that CENP-B and Sap1 would co-localize to the regions where they acted on the replication fork, and that these regions would engage in homologous recombination and degrade to double-strand breaks in the absence of CENP-B. To test this hypothesis, we performed chromatin immunoprecipitation of Sap1, Abp1 and Cbh1 followed by high-throughput sequencing (ChIP-seq). Abp1 has previously been shown to localize and recruit Cbh1 to the LTRs of Tf1 and Tf2 retrotransposons, where Abp1/Cbh1 play a role in their transcriptional silencing[2]. We demonstrated a strong co-localization of Sap1 with Abp1 and Cbh1 at these LTRs as well as at solo LTRs scattered throughout the genome (Fig. 2a, c and Supplementary Fig. 7a, b) and at the mating type locus (Supplementary Fig. 8), where Sap1 and Abp1 have been described to regulate mating-type switching[14,15]. Both Sap1 and Abp1/Cbh1 also localized to genomic regions independently of each other, suggesting that they do not form a stable complex or mediate their mutual recruitment. In particular, Abp1 exhibited binding to transfer RNA (tRNA) genes (Fig. 2b and Supplementary Fig. 7b), known to be potent replication pause sites[13,16]. Abp1 and Cbh1 co-localize to a highly A/T-rich region located in positions 100–150 of the LTR (Fig. 2c and Supplementary Fig. 7a, b). The localization of Sap1 within the LTR was concentrated in the first 50 base pairs of sequence (Fig. 2c), coinciding with a predicted Sap1-binding site[17] (Supplementary Fig. 7a, c). We tested this sequence by electrophoretic mobility shift assay and detected specific binding in wild-type extracts (Fig. 2d) as well as decreased binding and altered mobility in extracts from *Δabp1Δcbh1sap1-c* mutants (Supplementary Fig. 7d). Interestingly, solo LTR and full-length Tf2 insertions were associated with a prominent peak of Sap1 binding located outside the 3′ end of the transposon sequence (Fig. 2c). These observations indicate that Sap1 binding precedes and possibly guides Tf element integration. To test this prediction, we plotted the average enrichment of Sap1, Abp1 and Cbh1 around more than 70,000 *de novo* Tf1 integration sites recently reported[18]. We observed a dramatic association of these integration sites with a peak of Sap1 binding immediately downstream of the insertion site (Fig. 2e and Supplementary Fig. 8) and no appreciable CENP-B enrichment. These results strongly suggest that Sap1-binding sequences determine the targeting and orientation of Tf retroelement transposition.

To evaluate the mutual influence of Sap1 and Abp1/Cbh1 on LTR binding, we performed ChIP analysis of Sap1 in *Δabp1Δcbh1* mutants and of Abp1 in temperature-sensitive *sap1* mutants that affect DNA-binding activity[12]. Sap1 binding to the LTR was unaffected in *Δabp1* and *Δcbh1* mutants, and was slightly increased in *Δabp1Δcbh1* double mutants (Fig. 2f), but consistently reduced (twofold) in *Δabp1Δcbh1sap1-c*. Conversely, Abp1 binding to the LTR was increased between two and three times at the permissive temperature in *sap1-1* and *sap1-48* mutants (Fig. 2g). These results indicate that Sap1 and Abp1/Cbh1 bind to the LTR independently of each other and mutually counteract their recruitment, and that the *sap1-c* mutation impairs its binding to the LTR *in vivo* as well as *in vitro* (Supplementary Fig. 4).

A failure of replication-fork stability at LTRs, which are distributed throughout the genome, would explain the widespread DNA damage in *Δabp1Δcbh1* mutants. We assessed the behaviour of the replication
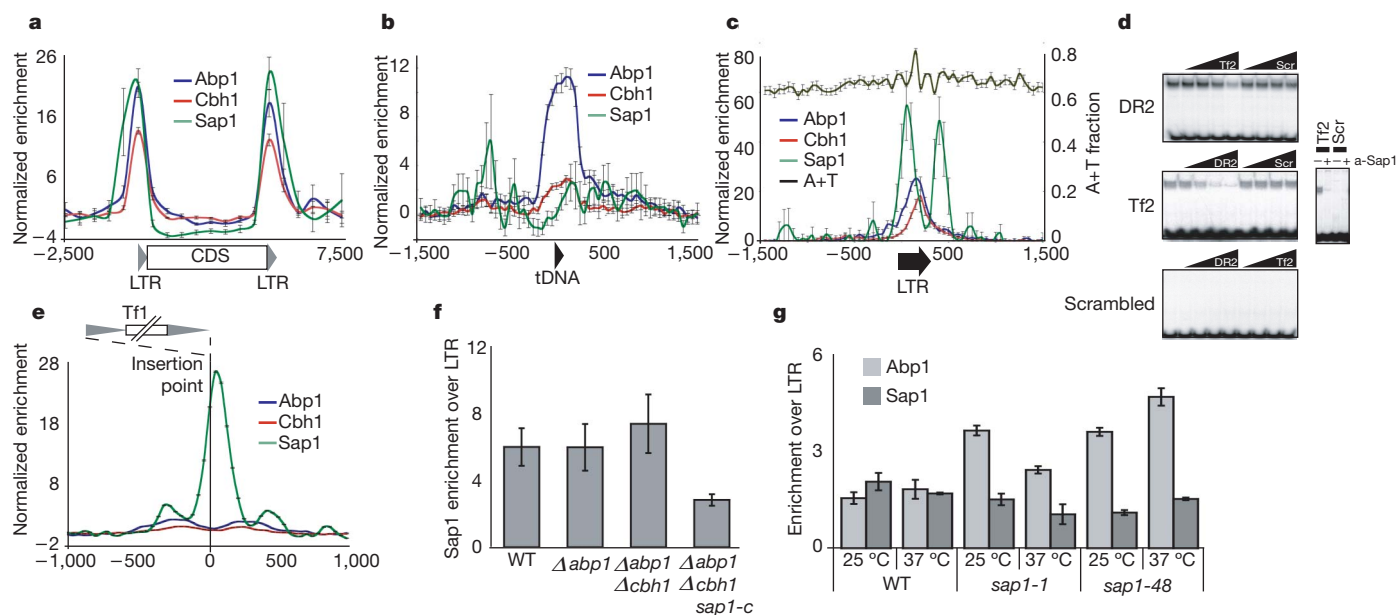

Figure 2 | Sap1 and CENP-B co-localize at the LTR of retrotransposons *in vivo*. Average genome-wide enrichment by ChIP-seq of Sap1, Abp1 and Cbh1 on (**a**) all *Tf2* elements, (**b**) euchromatic tRNA and (**c**) solo LTR. Error bars, s.e.m. **d**, Left panel, competition electrophoretic mobility shift assay; right panel, inactivation by incubation with anti-Sap1 serum[9]. **e**, Average Sap1, Abp1 and Cbh1 enrichment around Tf1 *de novo* insertion points[18]. **f**, ChIP of Sap1 with LTR of *Tf2* in CENP-B and *sap1-c* mutants and (**g**) of Abp1 with LTR of *Tf2* in *sap1* temperature-sensitive mutants. Error bars, s.d. for triplicates.
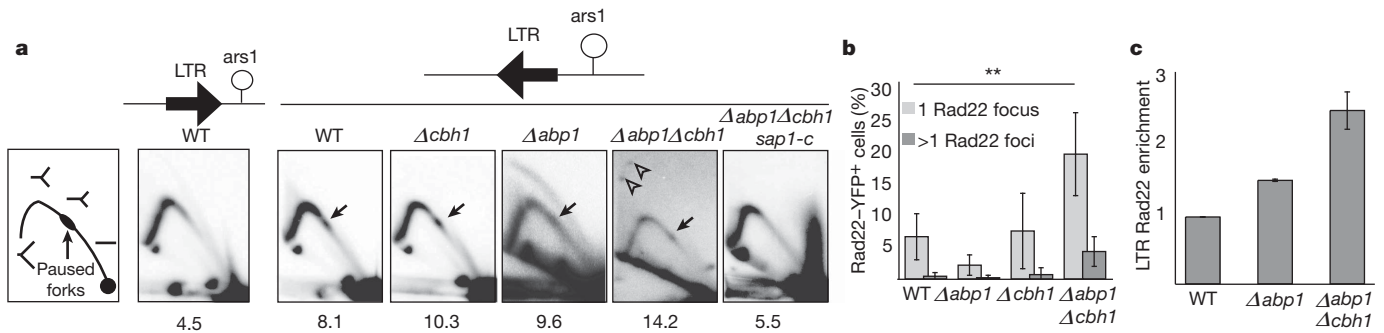
Figure 3 | CENP-B promotes replication-fork progression through the Sap1-dependent barrier present at the LTR and prevents homologous recombination. a, Two-dimensional gel electrophoresis of a plasmid fragment containing the *Tf2* LTR oriented towards (left) and away (right) from the *ars1* origin. Arrows, paused replication intermediates; open arrows, recombination intermediates. The percentage of signal over the LTR is indicated below each panel. **b**, Quantification of Rad22–YFP foci (*n* > 400 nuclei for all mutants). Error bars, s.e.m. **c**, Rad22–YFP ChIP with LTR in wild type, *Δabp1* and *Δabp1Δcbh1* mutants. Error bars, s.d. for triplicates.

fork as it traversed the LTR using two-dimensional agarose gel electrophoresis. Sap1-dependent programmed fork blocks are directional and only hinder fork progression in one orientation[10,11]. We cloned a full-length LTR and its first 50 base pairs (containing the Sap1-binding site) in a plasmid in both orientations with respect to the replication origin *ars1*. Two-dimensional gel electrophoresis in a wild-type strain transformed with this episomal system showed a modest accumulation of fork signal at the location of the cloned LTR (Fig. 3a), but only when the Sap1-binding site was proximal to the origin, and not in the opposite orientation (Supplementary Fig. 9). The Sap1-binding site was sufficient for this blocking activity, with the same orientation requirement (Supplementary Fig. 9). We next assayed the LTR for pausing activity in *Δabp1*, *Δcbh1* and *sap1-c* mutants (Fig. 3a). Strikingly, the paused fork signal was consistently enhanced and always at the same location in *Δabp1* and *Δcbh1* mutants, whereas the *Δabp1Δcbh1* double mutant exhibited additional signals outside the replication arc, suggestive of recombination intermediates[19]. The fork-blocking activity of the LTR disappeared in *Δabp1Δcbh1sap1-c* mutants. Unresolved fork blocks can collapse and undergo homologous recombination for fork recovery. We confirmed the presence of homologous recombination in the *Δabp1Δcbh1* double mutants by measuring the increase in the formation of Rad22 (homologous to Rad52 in *Saccharomyces cerevisiae*) foci in a Rad22–yellow fluorescent protein (YFP) strain[20] (Fig. 3b). We observed that *Δabp1Δcbh1* double-mutant cells accumulated the homologous recombination protein Rad22 at the LTR (Fig. 3c). Consistently, the recombination factor Rhp51 (Rad51 homologue) was essential for viability of *Δabp1Δcbh1* double mutants (Supplementary Fig. 10), indicating that homologous recombination is necessary for recovery from fork stalling at LTRs. These results indicate that Abp1/Cbh1 counteract Sap1 barrier activity and stabilize the replication fork at LTRs. This results in loss of genome integrity and homologous recombination at the LTR in *Δabp1Δcbh1* mutants.

The Sap1-binding sequence is conserved in Tf1 and Tf2 retrotransposon LTRs (Supplementary Fig. 7c), suggesting that it plays a role in the retrotransposon life cycle. We assayed the effect of *sap1* and *abp1/cbh1* on Tf2 stability by measuring the frequencies of loss of a *ura4* reporter transgene inserted in the Tf2-6 transposon[21]. Mutation of *abp1* resulted in a dramatic decrease of Tf2 ectopic recombination, which returned to normal levels when *sap1* was also mutated (Fig. 4). In the presence of *sap1+* there is a preference for gene conversion, which normally constitutes most ectopic recombination events[22]; however, in *Δabp1sap1-c* and *Δcbh1sap1-c* mutants the proportion of eviction and conversion events is similar (Fig. 4). Therefore we propose that the LTR recruits Sap1 to control the direction of transposon replication and increase transposon persistence in the genome, perhaps by coordinating lagging strand synthesis, which prevents single-strand annealing from complementary direct repeats (Supplementary Fig. 11a, b). CENP-B counteracts this activity, possibly by promoting replication-fork progression

through the Sap1-dependent barrier. Thus CENP-B and Sap1 promote genome and transposon integrity, respectively, in a 'tug-of-war' between transposon and host. Abp1 stimulates fork progression by recruiting the
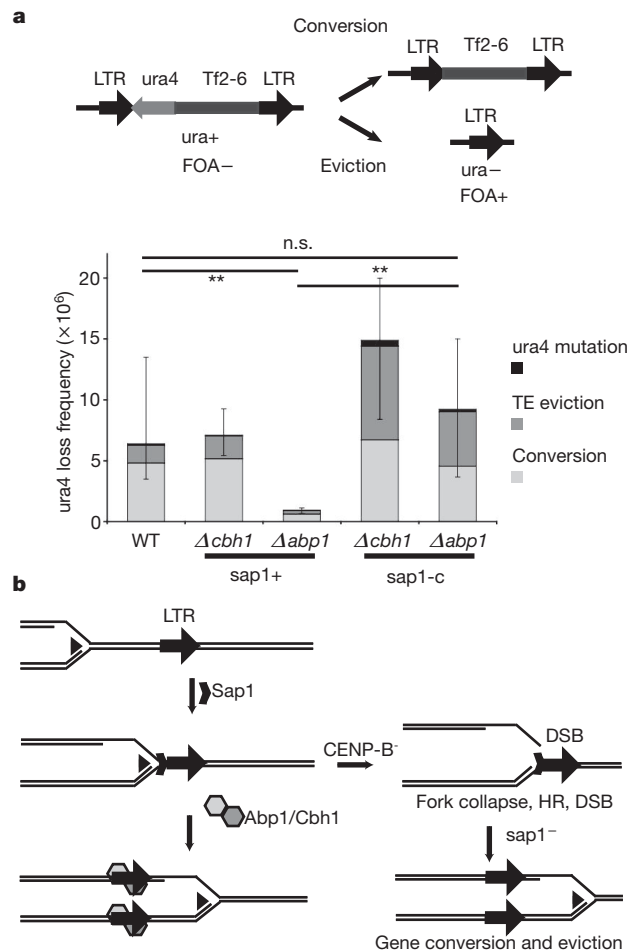

Figure 4 | CENP-B and Sap1 have opposite effects on Tf2 stability. a, Ectopic recombination fluctuation assay. Two potential mechanisms of *ura4* loss from the marked *Tf2-6::ura4* are indicated: gene conversion and eviction by LTR recombination. Columns represent total median *ura4* loss frequency in wild type, *Δabp1*, *Δcbh1*, *Δabp1sap1-c* and *Δcbh1sap1-c* mutants; error bars, 95% confidence intervals. Tints indicate distribution of mode of ectopic recombination events in the *ura4⁻* colonies obtained from wild type (*n* = 93), *Δabp1* (*n* = 88), *Δcbh1* (*n* = 94), *Δabp1sap1-c* (*n* = 91) and *Δcbh1sap1-c* (*n* = 89) mutants. **b**, Model for the interactions between Abp1, Cbh1, Sap1 and the replication fork at the LTR.

fork-restart protein MCM10 (ref. 4), which has primase activity. Additionally the histone deacetylase Mst1, which has roles in replication-fork stability interacts directly with Cbh1 (ref. 23). In *S. cerevisiae* the histone deacetylase Sir2 silences and inhibits recombination in repetitive DNA[24]. CENP-B factors recruit the histone deacetylases Clr3 and Clr6, which perform LTR silencing[2]. The result of these functions would be to preserve genome integrity at LTRs by preventing DNA damage and recombination. This novel role of CENP-B may not be limited to LTR and tDNA, as mutation of the replication-fork blocking factor *reb1*, which is specific to rDNA repeats, also suppresses the slow growth of the *abp1* mutant[25]. Similarly, our ChIP-seq data indicate that Sap1 may also be implicated in the functionality of the replication terminator RTS1 (Supplementary Fig. 8) in collaboration with Rtf1. In this manner, the function and regulation of the Sap1-bound regions is determined by the binding in their vicinity of different factors affecting replication-fork progression.

Because of their repetitive nature, transposons have a close relationship with replication and recombination. For example, the IS608 transposon of *Escherichia coli* is targeted to the lagging strand and always replicated in the same direction[26]. This might prevent recombination between tandemly arranged copies. We have shown that retrotransposons influence DNA replication through recruitment of directional fork blocking factor Sap1 and that activity of CENP-B is required for replication-fork management. Additionally, retrotransposition is targeted to the genomic localization of Sap1. These mechanisms influence the replicative dynamics of the host genome. The genomes of eukaryotes show widespread colonization by retrotransposons, and pericentromeric satellite repeats are often of transposon origin[27]. When such sequences are arranged as tandem repeats, control of replication direction by CENP-B would prevent chromosome breaks and preserve genome integrity. This mechanism accounts for the role of other regulators of fork progression in inter-LTR recombination[28,29]. In contrast, when flanked by LTR in opposite orientations, fragile sites fail to replicate and result in chromosome breaks[13,30].

METHODS SUMMARY

ChIP was performed using tagged TAP-Abp1 and TAP-Cbh1 strains with an Anti-Calmodulin Binding Protein antibody (Millipore) and a polyclonal serum against the native Sap1 protein[9]. High-throughput sequencing was performed on an Illumina G2 genome analyser, and analysed for polymorphism detection or statistical analysis of enrichment. Two-dimensional gel electrophoresis was performed as described[11]; see Supplementary Information for construction of the episomal system. Electrophoretic mobility shift assay was performed as described previously[17].

1. Okada, T. *et al.* CENP-B controls centromere formation depending on the chromatin context. *Cell* **131**, 1287–1300 (2007).
2. Cam, H. P., Noma, K., Ebina, H., Levin, H. L. & Grewal, S. I. Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature* **451**, 431–436 (2008).
3. Murakami, Y., Huberman, J. A. & Hurwitz, J. Identification, purification, and molecular cloning of autonomously replicating sequence-binding protein 1 from fission yeast *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **93**, 502–507 (1996).
4. Locovei, A. M., Spiga, M. G., Tanaka, K., Murakami, Y. & D'Urso, G. The CENP-B homolog, Abp1, interacts with the initiation protein Cdc23 (MCM10) and is required for efficient DNA replication in fission yeast. *Cell Div.* **1**, 27 (2006).
5. Lee, J. K., Huberman, J. A. & Hurwitz, J. Purification and characterization of a CENP-B homologue protein that binds to the centromeric K-type repeat DNA of *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **94**, 8427–8432 (1997).
6. Baum, M. & Clarke, L. Fission yeast homologs of human CENP-B have redundant functions affecting cell growth and chromosome segregation. *Mol. Cell. Biol.* **20**, 2852–2864 (2000).
7. Irelan, J. T., Gutkin, G. I. & Clarke, L. Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics* **157**, 1191–1203 (2001).
8. Irvine, D. V. *et al.* Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. *Genome Res.* **19**, 1077–1083 (2009).
9. de Lahondes, R., Ribes, V. & Arcangioli, B. Fission yeast Sap1 protein is essential for chromosome stability. *Eukaryot. Cell* **2**, 910–921 (2003).
10. Mejia-Ramirez, E., Sanchez-Gorostiaga, A., Krimer, D. B., Schvartzman, J. B. & Hernandez, P. The mating type switch-activating protein Sap1 is required for replication fork arrest at the rRNA genes of fission yeast. *Mol. Cell. Biol.* **25**, 8755–8761 (2005).
11. Krings, G. & Bastia, D. Sap1p binds to Ter1 at the ribosomal DNA of *Schizosaccharomyces pombe* and causes polar replication fork arrest. *J. Biol. Chem.* **280**, 39135–39142 (2005).
12. Noguchi, C. & Noguchi, E. Sap1 promotes the association of the replication fork protection complex with chromatin and is involved in the replication checkpoint in *Schizosaccharomyces pombe*. *Genetics* **175**, 553–566 (2007).
13. Szilard, R. K. *et al.* Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. *Nature Struct. Mol. Biol.* **17**, 299–305 (2010).
14. Aguilar-Arnal, L., Marsellach, F. X. & Azorin, F. The fission yeast homologue of CENP-B, Abp1, regulates directionality of mating-type switching. *EMBO J.* **27**, 1029–1038 (2008).
15. Arcangioli, B. & Klar, A. J. A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient mat1 switching in *Schizosaccharomyces pombe*. *EMBO J.* **10**, 3025–3032 (1991).
16. Deshpande, A. M. & Newlon, C. S. DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
17. Ghazvini, M., Ribes, V. & Arcangioli, B. The essential DNA-binding protein sap1 of *Schizosaccharomyces pombe* contains two independent oligomerization interfaces that dictate the relative orientation of the DNA-binding domain. *Mol. Cell. Biol.* **15**, 4939–4946 (1995).
18. Guo, Y. & Levin, H. L. High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. *Genome Res.* **20**, 239–248 (2010).
19. Segurado, M., Gomez, M. & Antequera, F. Increased recombination intermediates and homologous integration hot spots at DNA replication origins. *Mol. Cell* **10**, 907–916 (2002).
20. Meister, P. *et al.* Nuclear factories for signalling and repairing DNA double strand breaks in living fission yeast. *Nucleic Acids Res.* **31**, 5064–5073 (2003).
21. Sehgal, A., Lee, C. Y. & Espenshade, P. J. SREBP controls oxygen-dependent mobilization of retrotransposons in fission yeast. *PLoS Genet.* **3**, e131 (2007).
22. Kupiec, M. & Petes, T. D. Allelic and ectopic recombination between Ty elements in yeast. *Genetics* **119**, 549–559 (1988).
23. Gomez, E. B., Nugent, R. L., Laria, S. & Forsburg, S. L. *Schizosaccharomyces pombe* histone acetyltransferase Mst1 (KAT5) is an essential protein required for damage response and chromosome segregation. *Genetics* **179**, 757–771 (2008).
24. Huang, J. & Moazed, D. Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. *Genes Dev.* **17**, 2162–2176 (2003).
25. Roguev, A. *et al.* Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **322**, 405–410 (2008).
26. Ton-Hoang, B. *et al.* Single-stranded DNA transposition is coupled to host replication. *Cell* **142**, 398–408 (2010).
27. Wong, L. H. & Choo, K. H. Evolutionary dynamics of transposable elements at the centromere. *Trends Genet.* **20**, 611–616 (2004).
28. Putnam, C. D., Hayes, T. K. & Kolodner, R. D. Specific pathways prevent duplication-mediated genome rearrangements. *Nature* **460**, 984–989 (2009).
29. Wallis, J. W., Chrebet, G., Brodsky, G., Rolfe, M. & Rothstein, R. A hyper-recombination mutation in *S. cerevisiae* identifies a novel eukaryotic topoisomerase. *Cell* **58**, 409–419 (1989).
30. Lemoine, F. J., Degtyareva, N. P., Lobachev, K. & Petes, T. D. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **120**, 587–598 (2005).

Author Contributions M.Z., B.A. and R.A.M. designed the experiments presented and wrote the paper. M.Z. performed and analysed the experiments. M.W.V. provided bioinformatic analysis. D.G. and D.V.I. provided strains. S.W. and J.B. performed additional experiments.

LETTER

Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis

Mustafa Köksal[1], Yinghua Jin[2,3], Robert M. Coates[2], Rodney Croteau[4] & David W. Christianson[1]

With more than 55,000 members identified so far in all forms of life, the family of terpene or terpenoid natural products represents the epitome of molecular biodiversity. A well-known and important member of this family is the polycyclic diterpenoid Taxol (paclitaxel), which promotes tubulin polymerization[1] and shows remarkable efficacy in cancer chemotherapy[2]. The first committed step of Taxol biosynthesis in the Pacific yew (*Taxus brevifolia*)[3] is the cyclization of the linear isoprenoid substrate geranylgeranyl diphosphate (GGPP) to form taxa-4(5),11(12)diene[4], which is catalysed by taxadiene synthase[5]. The full-length form of this diterpene cyclase contains 862 residues, but a roughly 80-residue amino-terminal transit sequence is cleaved on maturation in plastids[6]. We now report the X-ray crystal structure of a truncation variant lacking the transit sequence and an additional 27 residues at the N terminus, hereafter designated TXS. Specifically, we have determined structures of TXS complexed with 13-aza-13,14-dihydrocopalyl diphosphate (1.82 Å resolution) and 2-fluorogeranylgeranyl diphosphate (2.25 Å resolution). The TXS structure reveals a modular assembly of three α-helical domains. The carboxy-terminal catalytic domain is a class I terpenoid cyclase, which binds and activates substrate GGPP with a three-metal ion cluster. The N-terminal domain and a third 'insertion' domain together adopt the fold of a vestigial class II terpenoid cyclase. A class II cyclase activates the isoprenoid substrate by protonation instead of ionization, and the TXS structure reveals a definitive connection between the two distinct cyclase classes in the evolution of terpenoid biosynthesis.

Although the first structures of C_{10} monoterpene[7], C_{15} sesquiterpene[8,9] and C_{30} triterpene[10] cyclases appeared several years ago, the structure of the 'missing link' in this series—a C_{20} diterpene cyclase—has been unknown until now. Plant diterpene cyclases such as taxadiene synthase are perhaps the most intriguing because they are the largest terpenoid cyclases (800–900 residues) and they are believed to be the most closely related to the ancestral plant terpenoid synthase[11,12]. Triple-domain plant diterpene synthases are believed to have evolved through the fusion of single-domain and double-domain bacterial diterpene cyclases, which in turn evolved from ancient progenitors[13].

The two distinct classes of terpenoid cyclase have unrelated protein folds and use different substrate activation mechanisms[14–17]. A class I terpenoid cyclase uses a trinuclear metal cluster liganded by conserved motifs **DDXXD** and **(N,D)DXX(S,T)XXXE** (bold indicates typical metal ligands) to trigger the ionization of the isoprenoid substrate diphosphate group, which generates a carbocation to initiate catalysis. A class II terpenoid cyclase initiates carbocation formation by general acid catalysis, using the 'middle' aspartic acid in a DXDD motif to protonate an isoprenoid double bond or oxirane moiety. Taxadiene synthase lacks a DXDD motif but contains conserved metal-binding motifs and requires Mg^{2+} for optimal catalytic activity[5], indicating that it functions as a class I terpenoid cyclase.

Expression and analysis of N-terminal truncation variants of taxadiene synthase revealed that deletions of 60 or 79 residues yield catalytically active proteins, whereas deletions of 93, 113 or 126 residues yield catalytically inactive proteins[18]. These results implicate the N-terminal segment D[80]DIPRLSANYHGDL[93] in catalysis. The N-terminal truncation variant lacking 60 residues has been studied with deuterated[18–21] and fluorinated[22] analogues of GGPP. These studies suggest a cyclization mechanism (Fig. 1) in which the diphosphate leaving group, the 14,15 π bond, and the 10,11 π bond of GGPP are optimally aligned for leaving-group departure with the formation of a verticillen-12-yl carbocation intermediate in the first step(s) of catalysis. Conformational inversion followed by 11α,7α-proton transfer and transannular B/C ring closure subsequently generates the taxen-4-yl carbocation, whose deprotonation yields taxa-4(5),11(12)-diene. The intramolecular proton transfer required to initiate transannular B/C ring closure occurs without the assistance of an enzyme-bound base[20]. The base mediating the final deprotonation step has not yet been identified.

The successful crystallization of TXS required co-crystallization with Mg^{2+} and either 13-aza-13,14-dihydrocopalyl diphosphate (ACP) or 2-fluorogeranylgeranyl diphosphate (FGP) (molecular structures are shown in Supplementary Fig. 1). Although not active, this truncation variant is exceptionally stable and is the only form examined that yielded satisfactory crystals. Surprisingly, TXS contains three α-helical domains and harbours the folds of both class I and class II terpenoid cyclases (Fig. 2); this structure is representative of nearly all diterpene cyclases. The C-terminal domain (S553–V862) has the class I terpenoid synthase fold that was first observed in farnesyl diphosphate synthase[23] and subsequently observed and designated the class I terpenoid synthase fold[8,14] in monoterpene and sesquiterpene cyclases[7–9]. This fold is also observed in geranylgeranyl diphosphate synthase[24], which generates the substrate for diterpene cyclases. The N-terminal domain of TXS (M107–I135 and S349–Q552) together with the 'insertion' domain[25] (S136–Y348) comprise the double α-barrel class II terpenoid synthase fold that was first observed in the triterpene cyclase squalene-hopene cyclase[10] and later observed in oxidosqualene cyclase[26]. TXS shares no significant overall amino-acid sequence identity with these triterpene cyclases.

Comparison of TXS with other terpenoid cyclases reveals that cyclase architecture is modular in nature and can consist of one, two or three domains (Fig. 2). Bacterial and fungal sesquiterpene cyclases are single-domain enzymes that adopt the class I terpenoid synthase fold; the first such enzymes to yield crystal structures were pentalenene synthase[8] and trichodiene synthase[27], respectively. Plant monoterpene and sesquiterpene cyclases generally contain two domains: the C-terminal domain adopts the class I terpenoid synthase fold, and the N-terminal domain adopts an unrelated α-helical fold that, as first noted by Wendt and Schulz[14], is homologous to the N-terminal domain of the class II triterpene cyclase squalene-hopene cyclase[10]. The first plant monoterpene and sesquiterpene synthases to yield crystal structures were bornyl diphosphate synthase[7] and 5-*epi*-aristolochene synthase[9], respectively. Most plant diterpene synthases contain three

[1]Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104-6323, USA. [2]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA. [4]Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340, USA.
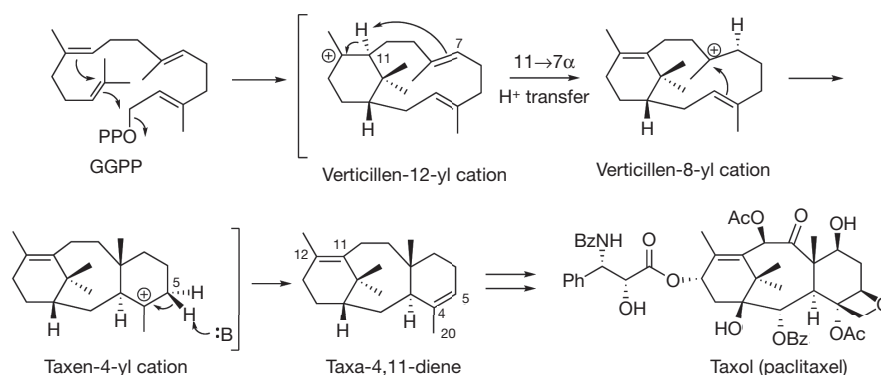
Figure 1 | Proposed catalytic mechanism of taxadiene synthase. The cyclization of GGPP to form taxadiene is the first committed step of Taxol (paclitaxel) biosynthesis in yew species. OPP, diphosphate; Ph, phenyl; Ac, acetyl; Bz, benzyl. Taxadiene is converted to Taxol through a lengthy series of oxidation and acylation steps.

domains, the third being an insertion conserved in sequence and position[25]. It was correctly predicted that this domain is homologous to the insertion domain of a triterpene cyclase on the basis of bioinformatics analysis[13].

It is interesting to note that the class II triterpene cyclases squalene-hopene cyclase[10] and oxidosqualene cyclase[26] are monotopic membrane proteins: each penetrates, but does not completely pass through, the membranes in which they are localized. Their triterpene substrates

(squalene and squalene oxide, respectively) are solubilized in the membrane and enter the active-site cavity through a hydrophobic channel open to the membrane surface. A nonpolar 'plateau' flanks the entrance to this channel near helix 8 in their respective insertion domains; helix 8 is quite hydrophobic in nature and probably serves as the membrane anchor (Fig. 2). In contrast, TXS functions in the plastid lumen, so its insertion domain does not contain the corresponding hydrophobic components.
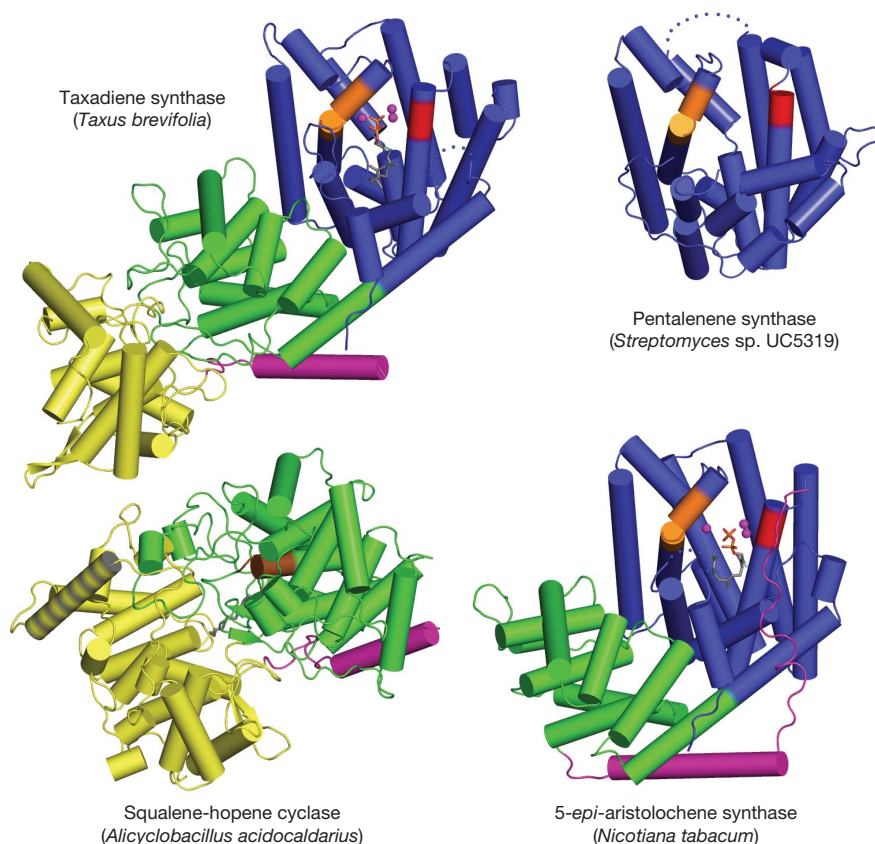

Figure 2 | Structural relationships among terpenoid cyclases. The class I terpenoid cyclase fold of pentalenene synthase[8] (PDB accession code 1PS1) (blue; 'α domain'[13]) contains metal-binding motifs DDXXD and (N,D)DXX(S,T)XXXE (red and orange, respectively); in 5-*epi*-aristolochene synthase[9] (PDB accession code 1LZ9) this domain is linked to a smaller vestigial domain (green; 'β domain'[13]). A related domain is found in the class II terpenoid cyclase fold of squalene-hopene cyclase[10] (PDB accession code 1SQC), where it contains the general acid motif DXDD (brown) and a second domain (yellow; 'γ domain'[13]) inserted between the first (purple) and second helices; a hydrophobic plateau flanking helix 8 (grey stripes) enables membrane insertion. Taxadiene synthase (PDB accession code 3P5R) contains both class I and class II terpenoid cyclase folds, but only the class I domain is catalytically active. The role of N termini (purple) in class I plant cyclases is to 'cap' the active site, as shown for 5-*epi*-aristolochene synthase.
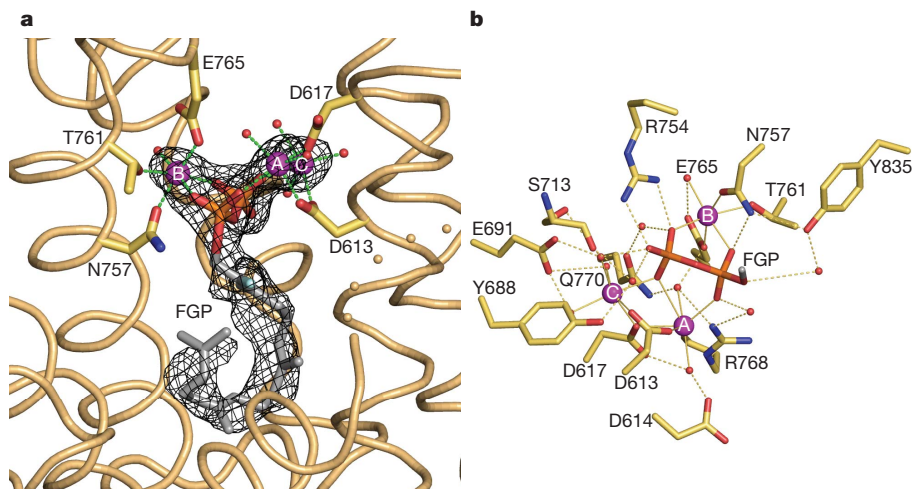
Figure 3 | Binding of substrate analogue to TXS. a, Simulated annealing $|F_o| - |F_c|$ omit map in which FGP and three Mg^{2+} ions are omitted from the structure factor calculation (contoured at 3.0σ); the side chains of metal ligands are indicated. **b**, Molecular recognition of the substrate diphosphate group in the TXS active site. For clarity, the isoprenoid moiety of FGP is truncated to one carbon (grey). Metal coordination and hydrogen bond interactions are indicated by thin solid and dashed lines, respectively. Atoms are colour-coded as follows: yellow, carbon; blue, nitrogen; red, oxygen; orange, phosphorus. Mg^{2+} ions (A, B and C) and water molecules are shown as purple and red spheres, respectively. A corresponding stereo figure is shown in Supplementary Fig. 3.

The active site of TXS is located in the C-terminal domain and is the exclusive binding site of the substrate analogue FGP (Fig. 3a and Supplementary Fig. 3a) and the bicyclic isoprenoid ACP (Supplementary Fig. 2; ACP does not mimic any intermediates in the TXS reaction, although it does mimic a common intermediate of many other diterpene cyclases). Metal-binding motifs that signal class I terpenoid cyclase function[15,27] are conserved in TXS as $D^{613}DMAD$ and $N^{757}DTKTYQAE$. The Mg^{2+}_A and Mg^{2+}_C ions are coordinated by D613 and D617, and the Mg^{2+}_B ion is chelated by N757, T761 and E765 (Fig. 3b and Supplementary Fig. 3b). Along with the recent observation of a trinuclear metal cluster in the active site of isoprene synthase[28], the structure of the TXS-Mg^{2+}_3-FGP complex indicates that three-metal ion catalysis is conserved across the greater family of class I terpenoid synthases: C_5 hemiterpene, C_{10} monoterpene, C_{15} sesquiterpene and C_{20} diterpene synthases.

In addition to metal coordination interactions, the diphosphate group of FGP also accepts hydrogen bonds from R754 and N757 (the latter residue also coordinates to Mg^{2+}_B) and makes water-mediated hydrogen bonds with Y688, E691, Y835, S713, R768 and Q770. It is interesting to compare the molecular recognition of the FGP diphosphate group with that of the product diphosphate group in the plant monoterpene cyclase bornyl diphosphate synthase[7] (Supplementary Fig. 3c). Most residues that assist the trinuclear metal cluster in binding and activating the substrate diphosphate group are conserved between these cyclases.

Class I terpenoid synthases undergo a significant structural transition from an open to a closed active-site conformation after the binding of three Mg^{2+} ions and the substrate diphosphate group, and this conformational transition helps to protect reactive carbocation intermediates from premature quenching by bulk solvent[15,16]. Although the structure of the fully open conformation of TXS is unavailable, we suggest that the structural changes observed between open and closed active-site conformations in plant monoterpene and sesquiterpene cyclases are representative of those that occur in the plant diterpene cyclase TXS. For example, active-site closure in bornyl diphosphate synthase[7] and 5-*epi*-aristolochene synthase[9] involves conformational changes of loops flanking the mouth of the active site; in addition, the N-terminal polypeptide 'caps' each active site. Specifically, the N-terminal polypeptide binds in a groove defined by the A–C and D–D1 loops on one side, and the J–K and H–H–α1 loops on the other. Tandem arginine residues in the N terminus of bornyl diphosphate synthase make key hydrogen-bond interactions in this groove. The N
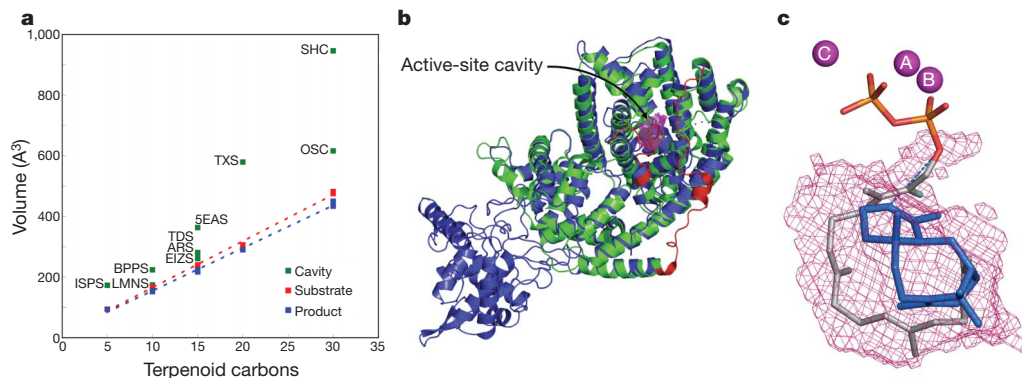

Figure 4 | Active-site cavities of terpenoid synthases. a, Active-site volumes are generally slightly larger than corresponding substrate and product volumes, perhaps to accommodate structural changes better during the cyclization cascade. Abbreviations are defined in Supplementary Table 2. **b**, Superposition of TXS (blue) and bornyl diphosphate synthase (green) guides the modelling of the J–K loop and the N-terminal segment of TXS (red) to define the enclosed active-site cavity (the magenta meshwork indicates the solvent-accessible surface). **c**, One orientation of taxadiene (blue) fits in the active-site cavity such that the H5β atom of the preceding taxen-4-yl cation would be oriented towards the diphosphate leaving group, suggesting that the PP$_i$ anion could serve as the stereospecific base that terminates the cyclization cascade. Three Mg^{2+} ions (A, B and C) and FGP are shown for reference; all protein atoms are omitted for clarity. A corresponding stereo figure is shown in Supplementary Fig. 4.

terminus of 5-*epi*-aristolochene synthase contains only a single corresponding arginine residue, R15, that seems to serve a similar function in the structure of the closed active-site conformation[9]. By analogy with the structures of these plant monoterpene and sesquiterpene cyclases, R84 in the missing N-terminal segment of TXS may help to stabilize the fully closed, catalytically active conformation of mature taxadiene synthase. Accordingly, the closed conformations of the J–K loop and the N-terminal segment of TXS are readily modelled on the basis of the bornyl diphosphate synthase structure to approximate the enclosed active-site contour that serves as the template for GGPP cyclization (Fig. 4 and Supplementary Fig. 4).

The active-site contour of TXS encloses a larger volume than the active sites of monoterpene or sesquiterpene cyclases, which is consistent with the larger isoprenoid substrate of the diterpene cyclase. The active-site cavity volumes of terpenoid synthases correlate with the hydrocarbon volume of their respective isoprenoid substrates (Fig. 4a and Supplementary Table 2). It has been suggested[16] that the shape of the active-site contour is more product-like for high-fidelity cyclases—that is, those that generate a single cyclization product—whereas if the active-site contour is less product-like, a more promiscuous cyclase results that generates multiple cyclization products. For TXS, the active-site volume is significantly larger than the volume of the product taxadiene. This is consistent with the observation that TXS is a somewhat promiscuous cyclase, generating about 20% of the alternative isomer taxa-4(20),11(12)-diene[21]. Indeed, the fact that TXS binds the bicyclic diterpene analogue ACP (Supplementary Fig. 2), which does not correspond to any intermediate in the TXS mechanism, clearly demonstrates promiscuity in ligand binding.

Taxadiene can fit in the enclosed active-site contour of TXS with two alternative orientations (Supplementary Fig. 4). Each orientation leads to possible suggestions for active-site bases that could function in the final deprotonation step of the cyclization cascade. Polar groups in the active site include S587, Q609, Y684, Y688, C719 and C830. Although one of these residues, for example Y688, could conceivably function as a base, taxadiene can also fit within the active-site contour such that H5β of the preceding taxen-4-yl carbocation would be oriented towards the inorganic pyrophosphate (PP$_i$) product (Fig. 4c and Supplementary Fig. 4). Thus, the PP$_i$ anion could serve as a stereospecific base, suggesting the possibility for substrate-assisted or product-assisted catalysis.

Finally, although the N-terminal domain and the insertion domain of TXS form a double α-barrel class II terpenoid synthase fold such as that characterizing the triterpene cyclases[10,26], the characteristic general acid DXDD motif and an active-site cavity are absent. Nevertheless, the TXS structure illuminates structure–function relationships in other diterpene cyclases that contain catalytically active class II cyclase domains. For example, consider the bifunctional diterpene cyclase abietadiene synthase from the grand fir tree (*Abies grandis*). Here, the class II terpenoid cyclase domain first catalyses the protonation-dependent cyclization of GGPP to form (+)-copalyl diphosphate, and the class I terpenoid cyclase domain then catalyses the ionization-dependent cyclization of (+)-copalyl diphosphate to form abietadiene[29]. Because the structures of abietadiene synthase and TXS are expected to be homologous, on the basis of 44% amino-acid sequence identity, the protonation-dependent reaction in the class II cyclase domain is presumably catalysed in much the same manner as for a triterpene cyclase reaction. In other diterpene cyclases such as copalyl diphosphate synthase from *Arabidopsis thaliana* (related to TXS by 31% amino-acid sequence identity), only the class II terpenoid cyclase domain is catalytically active; the class I terpenoid cyclase domain is vestigial and the signature metal-binding motifs are absent[30]. Thus, biosynthetic diversity in the family of terpenoid natural products is rooted in a 'mix and match' evolutionary strategy with class I and class II terpenoid cyclase folds, which can evolve together or separately as needed to generate the terpenoid product(s) required by the organism.

METHODS SUMMARY

A variety of different taxadiene synthase constructs were prepared, purified and assessed in crystallization trials, but only one proved satisfactory for crystallization. This construct, designated TXS, was one in which 107 residues were deleted from the N terminus and a hexahistidine tag was added to the C terminus to facilitate purification. TXS was expressed in *Escherichia coli* BL21 (DE3) RIL cells, purified, and co-crystallized with ACP or FGP by the sitting-drop vapour-diffusion method. The initial electron density map of the TXS–ACP complex was phased by using single-wavelength anomalous dispersion. After map fitting, refinement converged smoothly to $R/R_{free} = 0.167/0.205$. The structure of the TXS–FGP complex was solved by molecular replacement and refined to $R/R_{free} = 0.187/0.250$. Data collection and refinement statistics are shown in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Schiff, P. B., Fant, J. & Horwitz, S. B. Promotion of microtubule assembly *in vitro* by taxol. *Nature* **277,** 665–667 (1979).
2. Arbuck, S. G. & Blaylock, B. A. in *Taxol: Science and Applications* (ed. Suffness, M.) 379–415 (CRC Press, 1995).
3. Wani, M. C., Taylor, H. L., Wall, M., Coggon, P. & McPhail, A. T. Plant antitumor agents. VI. The isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **93,** 2325–2327 (1971).
4. Koepp, A. E. *et al.* Cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene is the committed step of Taxol biosynthesis in Pacific yew. *J. Biol. Chem.* **270,** 8686–8690 (1995).
5. Hezari, M., Lewis, N. G. & Croteau, R. Purification and characterization of taxa-4(5), 11(12)-diene synthase from Pacific yew (*Taxus brevifolia*) that catalyzes the first committed step of Taxol biosynthesis. *Arch. Biochem. Biophys.* **322,** 437–444 (1995).
6. Wildung, M. R. & Croteau, R. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of Taxol biosynthesis. *J. Biol. Chem.* **271,** 9201–9204 (1996).
7. Whittington, D. A. *et al.* Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl Acad. Sci. USA* **99,** 15375–15380 (2002).
8. Lesburg, C. A., Zhai, G., Cane, D. E. & Christianson, D. W. Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science* **277,** 1820–1824 (1997).
9. Starks, C. M., Back, K., Chappell, J. & Noel, J. P. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **277,** 1815–1820 (1997).
10. Wendt, K. U., Poralla, K. & Schulz, G. E. Structure and function of a squalene cyclase. *Science* **277,** 1811–1815 (1997).
11. Trapp, S. C. & Croteau, R. Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **158,** 811–832 (2001).
12. Keeling, C. I. *et al.* Identification and functional characterization of monofunctional *ent*-copalyl diphosphate and *ent*-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. *Plant Physiol.* **152,** 1197–1208 (2010).
13. Cao, R. *et al.* Diterpene cyclases and the nature of the isoprene fold. *Proteins Struct. Funct. Bioinf.* **78,** 2417–2432 (2010).
14. Wendt, K. U. & Schulz, G. E. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* **6,** 127–133 (1998).
15. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106,** 3412–3442 (2006).
16. Christianson, D. W. Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* **12,** 141–150 (2008).
17. Wendt, K. U., Schulz, G. E., Corey, E. J. & Liu, D. R. Enzyme mechanisms for polycyclic triterpene formation. *Angew. Chem. Int. Ed.* **39,** 2812–2833 (2000).
18. Lin, X., Hezari, M., Koepp, A. E., Floss, H. G. & Croteau, R. Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of Taxol biosynthesis in Pacific yew. *Biochemistry* **35,** 2968–2977 (1996).
19. Williams, D. C. *et al.* Heterologous expression and characterization of a 'pseudomature' form of taxadiene synthase involved in paclitaxel (Taxol) biosynthesis and evaluation of a potential intermediate and inhibitors of the multistep diterpene cyclization reaction. *Arch. Biochem. Biophys.* **379,** 137–146 (2000).
20. Williams, D. C. *et al.* Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase. *Chem. Biol.* **7,** 969–977 (2000).
21. Jin, Q., Williams, D. C., Hezari, M., Croteau, R. & Coates, R. M. Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labelling. *J. Org. Chem.* **70,** 4667–4675 (2005).
22. Jin, Y., Williams, D. C., Croteau, R. & Coates, R. M. Taxadiene synthase-catalyzed cyclization of 6-fluorogeranylgeranyl diphosphate to 7-fluoroverticillenes. *J. Am. Chem. Soc.* **127,** 7834–7842 (2005).
23. Tarshis, L. C., Yan, M., Poulter, C. D. & Sacchettini, J. C. Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-Å resolution. *Biochemistry* **33,** 10871–10877 (1994).

24. Chang, T. H., Guo, R. T., Ko, T. P., Wang, A. H. J. & Liang, P. H. Crystal structure of type-III geranylgeranyl pyrophosphate synthase from *Saccharomyces cerevisiae* and the mechanism of product chain length determination. *J. Biol. Chem.* **281,** 14991–15000 (2006).

25. Bohlmann, J., Meyer-Gauen, G. & Croteau, R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **95,** 4126–4133 (1998).

26. Thoma, R. *et al.* Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **432,** 118–122 (2004).

27. Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl Acad. Sci. USA* **98,** 13543–13548 (2001).

28. Köksal, M., Zimmer, I., Schnitzler, J.-P. & Christianson, D. W. Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. *J. Mol. Biol.* **402,** 363–373 (2010).

29. Peters, R. J., Ravn, M. M., Coates, R. M. & Croteau, R. Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* **123,** 8974–8978 (2001).

30. Prisic, S. & Peters, R. J. Synergistic substrate inhibition of *ent*-copalyl diphosphate synthase: a potential feed-forward inhibition mechanism limiting gibberellin metabolism. *Plant Physiol.* **144,** 445–454 (2007).

METHODS

Cloning, expression and purification of taxadiene synthase. Heterologous expression of taxadiene synthase from *Taxus brevifolia* lacking the N-terminal segment M1–V79 (M79-TXS) in *Escherichia coli* was achieved at the University of Pennsylvania, using procedures described previously[19]. We found that this protein consistently underwent degradation at 20 °C and 4 °C over a period of a few days to generate a soluble polypeptide stable for at least 4 weeks. Edman sequencing (Wistar Institute Proteomics Facility) showed that this polypeptide lacked the first 29 residues. Given its exceptional stability, this truncated polypeptide was considered a good candidate for crystallization. Accordingly, the M79-TXS gene segment corresponding to an N-terminal truncation at R107 (M107-TXS) was amplified by PCR with the following forward and reverse primers with flanking *Nde*I and *Bam*HI sites, respectively: 5′-GCACATATGGAGAGTTCTACT TACCAAGAAC-3′ and 5′-GCAGGATCCTACTTGAATTGGATCAATATAA AC-3′. A variant of the pET22b vector (pET22bTV; Novagen) was created by PCR with the following forward and reverse primers with complementary flanking restriction sites: 5′-GCAGGATCCCACCACCACCACCACCACC-3′ and 5′-GCACA TATGTATATCTCCTTCTTAAAGTTAAAC-3′. The gene encoding M107-TXS and the pET22bTV vector were ligated to generate a plasmid encoding the M107-TXS polypeptide with a C-terminal hexahistidine tag (M107-TXS-CHT), which was used to transform *E. coli* XL1Blue cells for amplification. The resulting clones were confirmed by DNA sequencing (University of Pennsylvania School of Medicine Sequencing Facility) to have only two silent mutations and no amino-acid substitutions.

The M107-TXS-CHT protein (henceforth designated 'TXS') was expressed in *E. coli* BL21 (DE3) RIL cells. Transformed cell cultures were grown in 2-l flasks containing 1 l of Luria–Bertani medium with 100 mg of ampicillin at 37 °C. At an attenuance (D_{600}) of 0.6–0.7, cultures were equilibrated at 20 °C and expression was induced by 0.25 mM isopropyl-1-thio-β-D-galactopyranoside for 16 h. Cells were harvested by centrifugation at 6,000*g* for 10 min, producing about 9 g of pellet per litre of culture. The pellet was suspended in 20 ml of buffer E (50 mM K_2HPO_4 pH 7.5, 300 mM NaCl, 10% (v/v) glycerol, 3 mM 2-mercaptoethanol) containing 1 mg ml^{-1} lysozyme and 1 mM phenylmethylsulfonyl fluoride, then incubated at 4 °C for 2 h with shaking. Cells were disrupted by sonication on ice six times (30 s on and 90 s off) with a large probe at medium power. Cell debris was cleared by centrifugation twice at 30,000*g* for 1 h. The clear supernatant was applied to a pre-equilibrated Talon column (Clontech Laboratories) at a flow rate of 1 ml min^{-1} with an ÄKTAprime plus fast performance liquid chromatography system (GE Healthcare Bio-Sciences AB). The loaded column was washed three times with 5 column volumes of buffer E, then buffer E plus 5 mM imidazole, then buffer E plus 10 mM imidazole. TXS was eluted with a gradient of 10–200 mM imidazole in buffer E at a flow rate of 2.5 ml min^{-1}. Selected fractions were combined, concentrated to a volume of 5 ml, and applied to a Superdex 200 preparatory-grade 26/60 size-exclusion column (GE Healthcare Bio-Sciences AB) with buffer A (25 mM 3-(*N*-morpholino)-2-hydroxypropanesulfonic acid (MOPSO) pH 6.8, 10% (v/v) glycerol, 1 mM dithiothreitol (DTT)) containing 300 mM NaCl. Fractions from this run were combined, concentrated to a volume of 5 ml and applied to the same column a second time with the same buffer. Fractions from the final size-exclusion column were combined and concentrated to 8.6 mg ml^{-1}. The purity of the TXS sample was 99% by SDS–PAGE analysis. No hexane-extractable products were identified by gas chromatography–mass spectrometry analysis after incubation with GGPP, indicating that this construct did not generate measurable amounts of taxadiene.

Crystallization. TXS could not be crystallized in the absence of isoprenoid diphosphate ligands. However, excellent crystals resulted when the protein was crystallized in the presence of ACP or FGP and Mg^{2+} ions by the sitting-drop vapour-diffusion method at 4 °C (ligand synthesis is outlined in Supplementary Information). To obtain the crystals of the TXS–ACP complex, a 1-μl drop of protein solution (5 mg ml^{-1} TXS, 25 mM MOPSO pH 6.8, 10% glycerol, 1 mM DTT, 2.5 mM ACP, 2.5 mM MgCl$_2$) was added to a 1-μl drop of precipitant solution (100 mM Bis-Tris pH 6.5, 25% polyethylene glycol 3350, 200 mM NaCl) and equilibrated against a 250-μl well reservoir of precipitant solution. Prism-like crystals with rounded edges appeared within 2–3 days and grew to maximal dimensions of 50 μm × 100 μm × 200 μm in 2–3 weeks. These crystals were flash-cooled after transfer to a cryoprotectant solution consisting of the mother liquor augmented with 15% ethylene glycol. For the preparation of a heavy-atom derivative for phasing, crystals of the TXS–ACP complex were soaked in a cryoprotectant solution (100 mM HEPES pH 7.5, 25% polyethylene glycol 3350, 100 mM NaCl, 100 mM MgCl$_2$, 10% glycerol) containing 2 mM methylmercury chloride for 22 h at 15 °C before flash-cooling.

To obtain crystals of the TXS–FGP complex, a 1-μl drop of protein solution (5 mg ml^{-1} TXS, 25 mM MOPSO pH 6.8, 10% glycerol, 1 mM DTT, 2.5 mM FGP, 2.5 mM MgCl$_2$] was added to a 1-μl drop of precipitant solution (100 mM HEPES pH 7.0, 20% polyethylene glycol 3350, 200 mM MgCl$_2$) and equilibrated against a 250-μl well reservoir of precipitant solution. These crystals were flash-cooled after transfer to a cryoprotectant solution consisting of the mother liquor augmented with 10% glycerol.

Collection and processing of X-ray diffraction data. Crystals of the TXS–ACP and TXS–FGP complexes diffracted X-rays to 1.82 Å and 2.25 Å resolution, respectively, at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, beamline X-29, using incident radiation with $\lambda = 0.945$ Å and 1.008 Å, respectively. Crystals of the mercury-derivatized TXS–ACP complex diffracted X-rays to 2.6 Å resolution at NSLS beamline X-25 using incident radiation with $\lambda = 1.000$ Å. All diffraction data were processed with HKL2000 (ref. 31). Crystals of the TXS–ACP complex belonged to space group $P2_12_12_1$ with unit cell parameters $a = 55.46$ Å, $b = 72.41$ Å, $c = 206.93$ Å, with one molecule in the asymmetric unit; the Matthews coefficient V_M was 2.35 Å^3 Da^{-1} (solvent content 48%). Crystals of the TXS–FGP complex belonged to space group $P2_1$ with unit cell parameters $a = 54.05$ Å, $b = 201.98$ Å, $c = 81.43$ Å, $\beta = 91.60°$, with two molecules in the asymmetric unit; $V_M = 2.61$ Å^3 Da^{-1} (solvent content 53%). Data collection and reduction statistics are shown in Supplementary Table 1.

Phasing and structure refinement. The initial electron density map of the TXS–ACP complex was phased by single-wavelength anomalous dispersion (SAD) with the 2.6-Å resolution data collected from the methylmercury chloride derivative. Initially, six Hg^{2+} atoms were located by using the program HKL2MAP[32] and used for SAD phasing, search and refinement of an additional seven Hg^{2+} sites; density modification, initial electron density map calculation and automatic model building were performed with the AUTOSOL routine implemented in PHENIX[33]. This procedure built more than 50% of the protein residues into the initial electron density map, most of which were α-helices. Manual model building subsequently generated an initial model with 90% of the residues registered in the sequence. This model was used for molecular replacement calculations with the AUTOMR routine implemented in PHENIX with the 1.82-Å resolution data collected from the TXS–ACP complex. Initial rigid-body refinement, iterative cycles of positional refinement, and grouped and individual atomic *B*-factor refinement were performed with PHENIX. Manual model rebuilding was performed with COOT[34]. Water molecules, Mg^{2+} ions and the ACP molecule were included in later cycles of refinement. A total of 745 out of 764 residues are present in the final model of the TXS–ACP complex; disordered segments excluded from the final model include N-terminal residues M107–S110 (M107 is the N terminus of the construct), the C-terminal hexahistidine tag and its associated linker residues (G863–H870) and surface loop I838–A844. An electron density map of the TXS–ACP complex is shown in Supplementary Fig. 2.

The model of the TXS–ACP complex without its ligand and solvent atoms was used as a search probe for molecular replacement calculations to solve the structure of the TXS–FGP complex at 2.25 Å resolution. Rigid-body refinement, positional refinement and grouped and individual atomic *B*-factor refinement were performed with PHENIX. Manual model rebuilding was performed with COOT. In the final model of the TXS–FGP complex, 746 and 736 out of 764 residues were present in monomers A and B, respectively. Disordered segments excluded from the final models of monomers A and B included N-terminal residues M107–S110, the C-terminal hexahistidine tag and its associated linker residues (G863–H870) and loop I574–R578; in addition, surface loop F837–E846 was disordered in monomer B.

For both structures, data reduction and refinement statistics are shown in Supplementary Table 1. Ramachandran plot statistics, calculated with PROCHECK[35], were as follows. TXS–ACP complex: allowed, 93.7%; additionally allowed, 5.9%; generously allowed, 0.3%; disallowed, 0.1%. TXS–FGP complex: allowed, 91.3%; additionally allowed, 8.1%; generously allowed, 0.4%; disallowed, 0.1%. Simulated-annealing omit maps were calculated with CNS[36]. Protein structure figures were prepared with the graphics program PyMol (http://www.pymol.org/).

Model of TXS in the fully closed conformation. To model the N terminus and J–K loop segments of TXS in a fully closed conformation and to calculate the active-site cavity volume, the N terminus (residues 54–81) and J–K loop (residues 574–587) segments of bornyl diphosphate synthase in its complex with three Mg^{2+} ions and 3-azageranyl diphosphate (PDB accession code 1N20) were 'grafted' onto the structure of the TXS-Mg^{2+}_3-FGP complex and mutated to the corresponding residues of TXS; S110 was also introduced to account for an insertion in the sequence alignment between TXS and bornyl diphosphate synthase. The conformations of the grafted segments were then subjected to 10,000 steps of gas-phase conjugate gradient energy minimization, using NAMD[37] and the CHARMM22 force field[38]. During energy minimization the grafted segments plus three adjacent residues on the N-terminal and C-terminal ends were unconstrained, while the remaining heavy atoms were fixed. Non-bonded cutoff and switch distances were set to 12 Å and 10 Å, respectively. The final structure resulting from this computation was used as the hypothetical fully closed conformation of TXS. The meshwork representing the active-site cavity of TXS was calculated

with VOIDOO[39], using a probe with a radius of 1.4 Å to generate a molecular surface based on solvent accessibility. To study product-binding orientations in the enclosed active site, a model of taxadiene was constructed on the basis of the coordinates of the taxane core of Taxol deposited in the Cambridge Crystallographic Data Centre with accession code TEYPAO[40].

The active-site cavity volume of TXS was compared with the active-site cavity volumes of other terpenoid cyclases, their substrates and their products (Fig. 4a and Supplementary Table 2). All volume calculations were performed with VOIDOO, using a probe with a radius of 0.0 Å to generate a molecular surface based on the atomic van der Waals radii. Because the active site of the hemiterpene synthase isoprene synthase was not fully closed as a result of disorder of the J–K loop, the active-site contour was artificially truncated by the placement of 'dummy' atoms to estimate the boundary of the fully enclosed cavity. We used a similar approach to model the cavity of TXS in a chemically sensible manner.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

32. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004).

33. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **D66**, 213–221 (2010).

34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr.* **D66**, 486–501 (2010).

35. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).

36. Brünger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921 (1998).

37. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

38. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

39. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr.* **D50**, 178–185 (1994).

40. Gao, Q. & Chen, S. H. An unprecedented side chain conformation of paclitaxel (Taxol®): crystal structure of 7-mesylpaclitaxel. *Tetrahedr. Lett.* **37**, 3425–3428 (1996).

CAREERS

R. MORRELL/CORBIS

CAREER PATHS

Where are they now?

Despite obstacles, many of those who kept Naturejobs *postdoc and graduate journals in the past six years have been successful — but not always in ways they had anticipated.*

BY PAUL SMAGLIK

Starting in 2004, *Naturejobs* began publishing accounts from graduate students of their hopes, frustrations, scientific victories and career defeats. We selected four journal keepers through a contest that attracted hundreds of applicants from dozens of countries. In 2007, we launched a similar postdoc journal. One writer called her scientific career a "winding road". Today, many of those writers would add that the road also presents potholes, detours and dead ends.

Career and life decisions may have roughened the road, but all are making their way — some taking less conventional routes than others. Of the 24 graduates and postdocs who kept journals between 2004 and 2009, we found the status of 21. Twelve — of whom five are still postdocs — are working as researchers at academic institutions or non-profit organizations. Of the other nine, three went into industry, one into government, three are engaged in science writing or communications and two left science altogether. But all shared a willingness to accept personal sacrifice and an ability to adapt to changing circumstances.

Despite the long odds, some former journal keepers have found academic positions. Tobias Langenhan, who wrote a graduate journal in 2005 while studying for a master's degree in neuroscience at the University of Oxford, UK, feels fortunate to have landed an independent position without doing a postdoc; he took a job at the University of Würzburg in Germany, where he went to medical school before his master's and PhD. But Langenhan says that he now faces a set of demands that he didn't anticipate when he started the journal. "The amount of time and effort that has gone into grant writing and planning out a scientific strategy for the next five years," he says, is something that he "definitely didn't expect back then".

Sidney Omelon wrote her graduate journal in 2004 while doing a biomedical ▶

engineering PhD at the University of Toronto in Canada. She was well aware of the poor prospects in academia — so much so that she made contingency plans for industry. But her 'plan A' panned out, with a job as an assistant professor of chemical and biological engineering at the University of Ottawa in Canada. Yet Omelon now worries about the personal sacrifices she has made along the way — ranging from affected relationships to a lapsed interest in rowing. "I worry that a combination of curiosity and stubbornness has brought my scientific career forward while leaving behind other aspects of my life," she says.

Maria Ocampo-Hafalla, who wrote a postdoc journal in 2007–08, is still working at Cancer Research UK in London. Her dreams of a university position remain alive, but she is approaching something of a crossroads as her five-year postdoc stint — including an extension for maternity leave — nears its end. Her husband is further along in his own scientific training, and Ocampo-Hafalla plans to use the Christmas holiday to ask herself some tough questions. Should she do another postdoc, in the hope of one day landing an academic position? If so, it will have to be in or near London, where her husband is now training. Or should she pursue another line of work, such as communication or policy? Ocampo-Hafalla isn't sure that she is ready to sacrifice family time to run experiments well into the night. "I don't necessarily want to be a lab rat, now that I'm a mother," she says. Still, it will be hard to let go of her research dreams. She has been doing bench work since she was 16.

INDUSTRY AIMS

One former journal writer who tested the academic waters didn't like what she found. "I was surprised at how much I disliked the academic interview process and the academic positions," says Moira Sheehan, who kept a journal in 2007–08 while a postdoc studying plant breeding at Cornell University in Ithaca, New York. She recalls one university where an interview involved repeated questions about family details, and another at which a lack of start-up funds and insufficient equipment and facilities would have quashed any chances of getting enough work published to gain a competitive tenured position in the time allotted by the university.

For Sheehan, industry was the right decision both personally and professionally, but it required some juggling. She is now a research scientist and project manager at Nature Source Genetics, a computational genomics company based in Ithaca, but while in the academic system she dealt with repeated funding shortfalls. When Sheehan's own grant ran out while she was a postdoc, her adviser supported her for eight months until he, too, fell short of money and had to let her go. She considered changing fields completely to get funding, but soon secured a research-associate position at

Cornell, focusing on maize research — a post that would let her complete a journal paper. Four months later, a grant she had written under her former adviser came through; he offered to take her back, but she stayed with her new lab.

"The hardest part of the postdoc was how it ended, and the spectre of unrealized potential I had in that lab, in that career path and field," says Sheehan, who wonders whether earlier publication of the journal paper might have earned her and her adviser a grant. "I never had a chance to make more out of it, which would have given me a better foundation for

Journal keepers: clockwise from top left, Moira Sheehan, Milan de Vries, Joanne Isaac, Tobias Langenhan, Sidney Omelon, Tshaka Cunningham.

the academic positions to which I applied," she says (for advice on navigating the quagmire, see 'Survival Tips').

Jason Underwood wrote for *Naturejobs* in 2005, as a PhD student in molecular biology at the University of California, Los Angeles. After graduating, travelling to Alaska and teaching at Occidental College in Los Angeles, he found a position at Pacific BioSciences, a start-up that makes next-generation genetic sequencers in Menlo Park, California. Underwood likes his current work–life balance — he and his family live only seven miles away from the company, making for an easy commute. And in industry,

he doesn't have to worry about funding or tenure. "I want to live a life — not just be grappling for grants," he says. He wouldn't mind doing more teaching — but he doesn't want to do a postdoc to earn that honour.

ALTERNATIVE ROUTES

Milan de Vries, who kept a graduate journal in 2006 while studying for a PhD in molecular biology at the Massachusetts Institute of Technology (MIT) in Cambridge, recalls how a random e-mail oversight during his first year as a postdoc at MIT added to his considerable anxiety about securing a university post. A postdoc colleague had applied for a faculty position at another research university and had received a form-letter e-mail rejection. Whoever had sent it had mistakenly selected 'cc' rather than 'bcc', so de Vries's colleague could see the names and institutions of the other spurned applicants — all 300 or so of them. "That's not a very encouraging result," says de Vries, who knew that even those who do land a job still have to secure funding and tenure.

Meanwhile, de Vries had been nurturing an interest in politics, through a group of friends who were supporting Barack Obama's US presidential campaign in 2008. That eventually led him to his current job as campaign manager at MoveOn.org, a grass-roots political action group. De Vries's position there has some parallels to a research job in hard science: both focus on databases and project management. "A database containing people's political info versus a database with DNA info is basically the same," he says.

Peter Jordan, who kept a postdoc journal in 2007 while at the US National Institute for Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland, made an even more extreme move: to theology. He had felt slightly unsatisfied with his PhD work on mathematical modelling of biological systems research — but couldn't put his finger on why, although he had a slowly growing interest in theology. That feeling increased during the first year of his postdoc, when he started writing for *Naturejobs*. Eventually, he left the postdoc for the divinity school at Duke University in Durham, North Carolina, where he is now in the last year of his master's study. "Science is only one of many activities in which human beings engage," says Jordan. "Don't let it rule your life, and don't think that it is the only plausible perspective on the world."

For Amber Jenkins, who wrote a graduate journal in 2004, science communication was the alternative destination — although she took a circuitous path. For her PhD at Imperial College, London, she conducted particle physics research at Fermilab in Batavia, Illinois. While writing up her doctoral thesis, she worked as maternity cover for a *Nature* manuscript editor. That led to a job launching *Nature Photonics*. Jenkins eventually moved to San Francisco, California, to be closer to the photonics

DAVID TAYLOR

research community and her now husband. After another move and some consulting and freelance work, she found a full-time communications position at NASA's Jet Propulsion Lab (JPL) in Pasadena, California.

Moves aside, Jenkins's career path has proved satisfying. At JPL, she spends much of her time sharing climate science with the public by managing and creating content for NASA's Climate Change website. She has also worked with NASA's Orbiting Carbon Observatory mission, which, when launched in 2013, will measure carbon dioxide levels in the atmosphere. Jenkins became passionate about communicating science halfway through her PhD. But the UK native never expected to end up in the United States; nor did she envisage switching fields several times, from particle physics to applied physics and photonics to climate change.

WORK–LIFE, UNBALANCED

Many former postdoc and graduate journal writers reported that career choices and dilemmas had adversely affected their personal lives. Off-kilter work–life balances strained marriages, complicated parental aspirations and stymied the pursuit of other interests. "It's a cliché, but striking that balance is not easy," says Zach Lippman, who kept a postdoc journal while studying agriculture at the Hebrew University of Jerusalem in 2008 and is now

an assistant professor of plant biology at Cold Spring Harbor Laboratory in New York. "That balance is different for everyone. I'm still learning that balance, and it requires adjustment frequently," he says.

Jenkins struggled with disparities between her own career and her husband's, but now they finally live in the same place: the San Francisco Bay Area in California. However, each has a long commute. "If you both have specialities, it's damned hard to find jobs in the same place," says Jenkins. Bouncing between jobs, continents and aspirations was sometimes tough, she says. "Having to accept that part of growing and changing and living your dreams means giving up on other dreams, and moving on from people you care about, was the toughest part of my training. Living thousands of miles from home was, at times, lonely and difficult."

For Langenhan, pursuing academia meant distance from his wife — who had an established business career in Germany — while he spent five years earning a PhD in Oxford. "It was tough," he says. "We travelled back and forth every two or three weeks." Their partnership survived the strain, but others were not so fortunate.

Tshaka Cunningham wrote for *Naturejobs* in 2004, while a PhD student in virology at Rockefeller University in New York. He moved to Paris in 2005 for a fellowship at the Pasteur Institute, which put pressure on his

first marriage. It eventually collapsed. He took another fellowship, at the US National Institutes of Health in Bethesda, Maryland, and married again in 2008. Wanting to avoid moving his family around, he switched to the other side of the bench as a scientific programme manager for the US Department of Veterans Affairs in nearby Washington DC. Although he misses bench work, Cunningham is still involved in academia through an adjunct position teaching molecular biology at Howard University in Washington DC.

"The balance is different for everyone. I'm still learning, and it requires adjustment frequently."

Motherhood delayed some careers, and permanently altered others. Sheehan faced the daunting challenge of having a second child mid-postdoc. She took three months off before returning to the bench. "I always felt people were watching me closely to see how I handled it all, and it was hard to forgive myself for being human," she says. She first worried about handling a project — devising a novel system for culturing meiocytes in maize — that others had failed to complete. Would she, too, fall short? Then, when she cracked it, she thought that her answer couldn't be correct, because it had come too easily. "It was very hard to think clearly as I was always sleep-deprived," says Sheehan, adding that her current position in industry has freed her from constant funding worries and helped her carve out more family time.

Joanne Isaac started writing her journal in 2009 as a postdoc in climate change and biodiversity at James Cook University in Townsville, Australia, but gave up a full-time research job to focus on being a parent. With their toddler in tow, Isaac followed her husband to California, Colorado and, finally, back to Australia while he performed fieldwork for his own postdoc. Isaac's feminist sensibilities initially bristled at the arrangement, which ran counter to her career goals when she first started as a journal keeper. But now she has made peace with it. Although she works as a freelance writer, Isaac's main focus is on motherhood.

"I learnt how to be a stay-at-home mum," she says. "Not as easy as you'd think after being a working mum."

Despite the personal and professional challenges most journal keepers face, none regrets the journey thanks to its intellectual rewards. Even Jordan, who left science for theology, says he learnt a great deal along the way.

"On the one hand, I feel like my scientific training was a massive detour and took up a lot of time that I cannot get back," he says. "On the other hand, I've no doubt that things would have turned out very differently had I not taken the path through science research." ∎

Paul Smaglik *is a freelance writer in Milwaukee, Wisconsin.*

SURVIVAL TIPS
Journal keepers' keys to success

● Many more doors can and will open to you than you can probably conceive of at the moment — some of them after graduate school.

● Don't lose sight of why you are pursuing a PhD or postdoc. You're not there to be a good analyser of a particular sequence of DNA or a good worker. You're there to be a good scientist.

● Present yourself confidently. Establish a trustworthy and nurturing professional network. Value your time, and use it accordingly. Maintain your sense of wonder, humour and self.

● Science is only one of many activities. Don't let its practice rule your life.

● Remember why you are doing this, so that during down times you can remind yourself of your goal. It may cheer you up and, ideally, give you motivation to continue.

● Explore career options outside academia. If a faculty job is your goal, have at least one alternative. Discuss your career plans with others.

● Study and explore what truly interests you, even if you think it might hold you back from the perfect job.

● Stay open to new ideas and potential futures. World markets can change rapidly, and can often negate a career path.

● Know your personality and have a keen grasp of your strengths, weaknesses and personal and professional goals when deciding on your best career option.

● Your supervisor can promote or ruin your career. You need him or her to write a good recommendation. Be sure that you trust your supervisor before agreeing to work in his or her lab; try talking to the lab's past graduate students and postdocs.

● It is probably easiest to get a job by doing a postdoc with a well-known, established researcher. Such a position, however, may make it more difficult to pursue truly innovative work.

● Observe the management system within which you are working, and how the systems and people around you operate.

● Never underestimate how competitive and political people can be.

● Turn on your swagger. Beware of sharks. If you're self-effacing or modest, train in assertiveness or take up kick-boxing. **P.S.**

THE LAST LABORATORY

Raise a glass to chemistry.

BY JOHN GILBEY

From the window of our coffee room you get an excellent view of the Gothic, vaguely sinister, Registry building. My colleagues and I, the hardcore survivors of the New Year lunch party, watched with a sense of numbed expectation as two porters dismantled the Christmas tree in the dank, grey twilight. January is not the right time for bad news.

The flag hanging limply above the door showed that the Senate was in session, and, although the agenda used language like 'Management of Academic Risk Resulting From Tactical Resource-Balancing', it was clear what the real decision was expected to be. After 100 years of solid, if unspectacular, teaching and research, the University of Rural England intended to close its chemistry department.

Our professor strode from the gaunt limestone portal, paused, then threw his meeting papers to the ground. The porters looked on in bemusement as he began, slowly at first and then with uncontrolled fury, to jump up and down on the documents — mashing them to pulp in the melting snow. As our phones pinged with the expected incoming e-mail, we turned dejectedly from the window and raised our glasses in a solemn toast to the department. The party broke up in dissolute disillusion, but as I put on my coat, I realized that one important person would still be ignorant of the news, so I went to find Jim — our glass-blower, tea-brewer and father-confessor.

I stopped in the open doorway and tapped gently on the frame. "With you in a minute, John …" came a slightly muffled voice from behind a stereo microscope. As the gas jet roared, I looked around the cluttered space that was Jim's workshop. Old bits of glassware, tagged like the occupants of a morgue, stood in heaps; colourful family photos — tokens of thanks from former students — covered one noticeboard; and grey industrial ovens throbbed in the shadows.

The elderly man sat back from the dark teak bench, stretched mightily and turned to look at me. "They've decided then," he stated, waving me towards an old horsehair sofa half covered in junk. I nodded and settled into it — shoving aside a pile of old journals.

"How did you know? You don't use e-mail, do you?"

Jim laughed and pushed his safety glasses farther up his nose. "Nah, I don't hold with computers — they make life too complicated. Sid the porter was just in, told me about the prof's little dance routine over the road — he reckoned he might show some promise if he ever took up the tango."

I tried to picture this development — but failed.

"Anyway," said Jim. "What are they going to do? You can't just stop teaching chemistry …"

I explained the plan to form an integrated science department, building on what physics and computer science had already
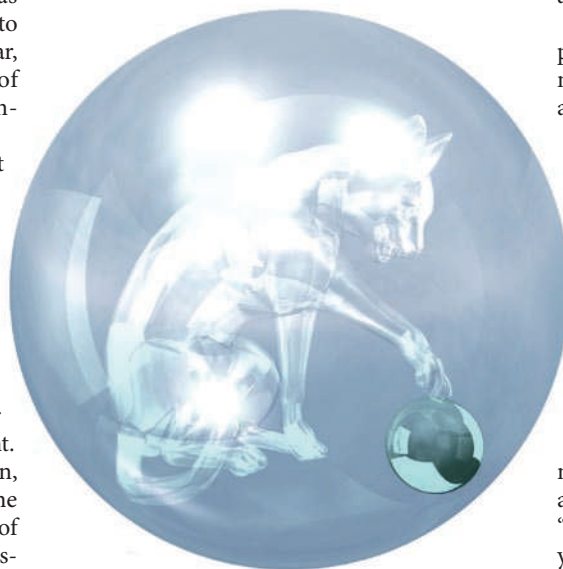
attempted. He shook his head as I told him about the closure of the labs themselves — with all practical study being replaced by simulations, and the generation of materials being restricted to the new, robot-driven Synthetics Fabrication Centre on the edge of the campus.

Jim scratched his bald head. "Doesn't seem right to me somehow," he said. "People need to see, touch and smell the stuff they're working with." I pointed out that the new VR system did exactly that, but Jim waved the concept away. "It just ain't the same — look …" He reached across and handed me a glass ball about a centimetre across. It felt cool and light, with an exquisite silky texture

— and when I gazed into it I could see a tiny glass figure, a tabby kitten frozen in the act of dabbing at a coloured ball. It was utterly perfect.

"That's amazing — how did you do it?"

The glass-blower allowed himself a slow smile. "That would be telling … It's for young Mercedes — she's got her viva next week, and I reckon she'll sail through. Here, take a closer look." He handed me an unmounted glass lens and I held it in front of the ball.

"Have you got anything with a bit more magnification?" I asked. "This doesn't really help much."

Jim leaned across conspiratorially, "Give it a squeeze …" he said with a grin.

I did, and nearly dropped it. As I applied pressure to the cold glass lens, it deformed in my grip — remaining circular but shrinking and becoming thicker.

The lens now made the kitten look close to life size — and it still looked perfect. I tried a couple of questions, but couldn't get beyond "How …?"

Jim eased both lens and ball from my fingers and placed them carefully back on the bench, then perched on the arm of the sofa. "I was apprenticed to Mr Chambers — my predecessor — for seven years, back in the day. He had some big ideas about the chemistry of glass — it was him got me into using botanicals." Jim nodded to a shelf full of dark bottles I'd long assumed to contain home-brewed wine. "Folk have been doing that for thousands of years — adding wood ash and such to glass. I just happened on some especially useful tinctures. That's what chemists are supposed to do, isn't it?"

"But Jim, this changes everything!" I whooped. "They can't close us down now!" A sudden thought struck me. "Will the lenses be easy to manufacture?"

Jim looked suddenly glum. "Ah, there's the thing. It's really tricky — all to do with balancing internal stresses. The secret's in the annealing. It's certainly nothing those fancy 3-D printers could handle. I'll have to train folk up to do it."

In my mind, I tried to picture what sort of dance the professor would attempt this time. ∎

John Gilbey *is visionary-in-residence at the University of Rural England. He is proud to have been taught by many intriguing technical staff — all of whom had a few tricks up their sleeves.*

2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

LETTER

c–Jun N–terminal phosphorylation antagonises recruitment of the Mbd3/NuRD repressor complex

Cristina Aguilera[1]*, Kentaro Nakagawa[1,2]*, Rocio Sancho[1], Atanu Chakraborty[1], Brian Hendrich[3] & Axel Behrens[1]

AP-1 (activator protein 1) activity is strongly induced in response to numerous signals, including growth factors, cytokines and extracellular stresses[1]. The proto-oncoprotein c-Jun belongs to the AP-1 group of transcription factors and it is a crucial regulator of intestinal progenitor proliferation and tumorigenesis[2–4]. An important mechanism of AP-1 stimulation is phosphorylation of c-Jun by the Jun amino-terminal kinases (JNKs)[1]. N-terminal phosphorylation of the c-Jun transactivation domain increases target gene transcription[5,6], but a molecular explanation was elusive. Here we show that unphosphorylated, but not N-terminally phosphorylated c-Jun, interacts with Mbd3 and thereby recruits the nucleosome remodelling and histone deacetylation (NuRD) repressor complex. Mbd3 depletion in colon cancer cells increased histone acetylation at AP-1-dependent promoters, which resulted in increased target gene expression. The intestinal stem cell marker *lgr5* was identified as a novel target gene controlled by c-Jun/Mbd3. Gut-specific conditional deletion of *mbd3* (*mbd3*[ΔG/ΔG] mice) stimulated c-Jun activity and increased progenitor cell proliferation. In response to inflammation, *mdb3* deficiency resulted in colonic hyperproliferation and *mbd3*[ΔG/ΔG] mice showed markedly increased susceptibility to colitis-induced tumorigenesis. Notably, concomitant inactivation of a single allele of *c-jun* reverted physiological and pathological hyperproliferation, as well as the increased tumorigenesis in *mbd3*[ΔG/ΔG] mice. Thus the transactivation domain of c-Jun recruits Mbd3/NuRD to AP-1 target genes to mediate gene repression, and this repression is relieved by JNK-mediated c-Jun N-terminal phosphorylation.

We described a yeast three-hybrid approach to identify proteins that interact with c-Jun depending on its phosphorylation status previously[7]. Mbd3 was cloned as an interactor that specifically bound to unphosphorylated c-Jun. Mbd3, together with Mbd1, 2 and 4, was originally characterized as a protein containing a region with high homology to the methyl-CpG-binding domain (MBD) of MeCP2. Mbd2 and Mbd3 assemble into mutually exclusive distinct NuRD complexes[8]. NuRD mediates gene repression through histone deacetylation via HDAC1 and 2, and chromatin remodelling ATPase activities through its CHD3 (Mi2α) and CHD4 (Mi2β) subunits[9–11]. Whereas Mbd2 recruits NuRD to methylated DNA, the MBD of Mbd3 fails to bind methylated DNA[12,13]. Thus, how Mbd3 interacts with chromatin to regulate transcription was not known.

Overexpressed c-Jun co-immunoprecipitated with Mbd3 but JNK activation by anisomycin (Ans) reduced the interaction of wild-type c-Jun and Mbd3. In contrast, a c-Jun4A mutant, in which the four N-terminal residues phosphorylated by JNK, serines 63 and 73, and threonines 91 and 93, are mutated into alanines, remained bound to Mbd3 despite high JNK activity (Fig. 1a). c-Jun4A also interacted more efficiently with CHD3 and MTA3, two major components of the NuRD complex (Supplementary Fig. 1a). Interaction of endogenous Mbd3 with c-Jun could not be detected when JNK activity was stimulated

by ultraviolet light, but pharmacological JNK inhibition (JNKi), which resulted in a significant decrease in c-Jun N-terminal phosphorylation (Supplementary Fig. 2a), stimulated Mbd3 binding to c-Jun (Supplementary Fig. 1b). Moreover, recombinant glutathione-S-transferase
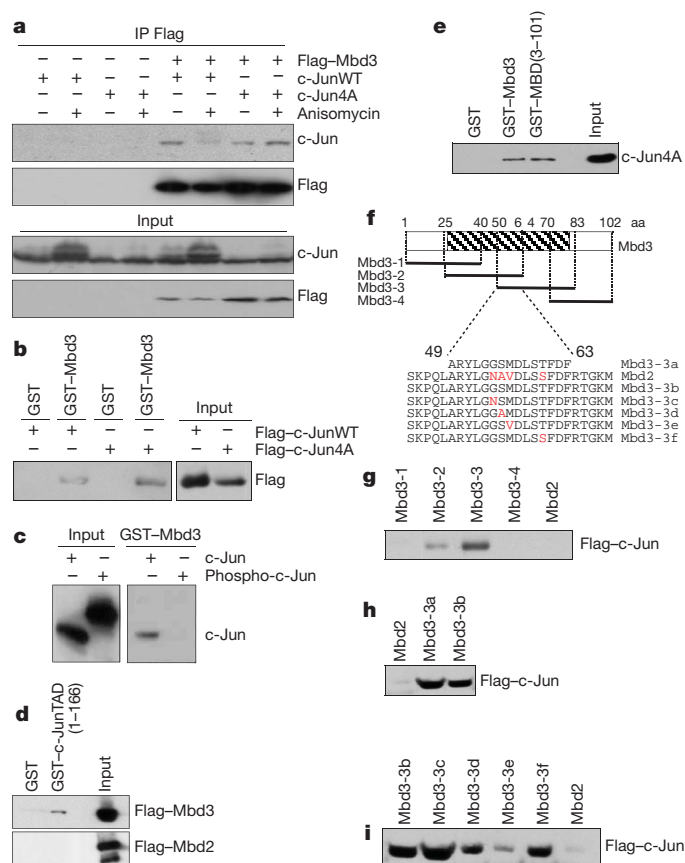

Figure 1 | Mbd3 interacts with unphosphorylated c-Jun through its MBD.
a, HCT116 cells transfected with Myc–c-JunWT or Myc–c-Jun4A along with Flag–Mbd3 were treated with anisomycin for 1 h where indicated. Cell lysates were used to immunoprecipitate (IP) Flag–Mbd3. **b**, Pull-down experiment using GST–Mbd3 with HCT116 cell lysates transfected with c-JunWT or c-Jun4A. **c**, Pull-down experiment using GST–Mbd3 as a bait with recombinant c-Jun (c-Jun) or recombinant c-Jun that was phosphorylated *in vitro* by incubation with activated JNK1 protein (Phospho-c-Jun). **d**, Pull-down experiment using GST–JunTAD(1–166) with HCT116 cell lysates transfected with Flag–Mbd3 or Flag–Mbd2. **e**, Pull-down experiment using GST–Mbd3 or GST–MBD with HCT116 cell lysates transfected with Myc–c-Jun4A. **f**, Schematic overview of Mbd3 peptides covering the MBD, mutations are highlighted in red. **g–i**, Pull-down experiments using different biotinylated peptides shown in **f** with HEK293T cell lysates expressing Flag–c-Jun.

[1]Mammalian Genetics Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. [2]Department of Medical Biochemistry, Graduate School of Medicine, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan. [3]Wellcome Trust Centre for Stem Cell Research, Medical Research Council Centre for Stem Cell Biology and Regenerative Medicine, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.
*These authors contributed equally to this work.

(GST)–Mbd3 interacted more efficiently with c-Jun4A than with wild-type c-Jun (Fig. 1b). Bacterially produced GST–Mbd3 bound to recombinant c-Jun, but this interaction was reduced by *in vitro* phosphorylation of recombinant c-Jun by activated JNK1 (Fig. 1c). GST pull-down assays showed that the transactivation domain (TAD, amino acid residues 1–166) of c-Jun was sufficient to bind Mbd3, but did not interact with the closely related Mbd2 protein (Fig. 1d). The delta (δ) domain of c-Jun (residues 31–60), which is required for JNK binding, was dispensable for Mbd3 interaction, and further mapping identified the region between residues 60 and 120, which encompasses the JNK phosphoacceptor residues, as the Mbd3 binding site of c-Jun (Supplementary Fig. 1c–e). Thus, Mbd3 binds directly and preferentially to the unphosphorylated TAD of c-Jun, and this interaction is disrupted by phosphorylation.

To determine the domain of Mbd3 involved in binding to c-Jun, serial deletion mutants of Mbd3 were generated (Supplementary Fig. 3a). The deletion of residues 40 to 80 of Mbd3, which encode part of the MBD, reduced Mbd3 binding to c-Jun (Supplementary Fig. 3b). Recombinant Mbd3 MBD (GST–MBD(3–101)) bound to c-Jun as efficiently as full-length Mbd3 (GST–Mbd3) (Fig. 1e). Using overlapping biotinylated peptides covering the Mbd3 MBD we identified the region from residue 49 to residue 63 as crucial for Mbd3 binding to c-Jun (Fig. 1f, g and Supplementary Fig. 4a). A 15-residue peptide covering this region was sufficient to mediate c-Jun interaction (Fig. 1h). This region differs only in four residues to the same region in Mbd2 and a peptide with the corresponding Mbd2 sequence did not bind c-Jun (Fig. 1h). Single amino acid changes revealed a large contribution of methionine 56 and to a lesser extent serine 55 of Mbd3 in c-Jun binding (Fig. 1i and Supplementary Fig. 4b). Thus, the MBD of Mbd3 is necessary and sufficient for c-Jun binding, indicating that the MBD of Mbd3 has evolved from a methyl-CpG-binding domain into a protein–protein interaction module.

To investigate the role of Mbd3 in regulating c-Jun function, we knocked-down *mbd3* with a small hairpin RNA (shRNA) construct (Supplementary Fig. 2b) and transfected it along with a luciferase reporter construct driven by an artificial promoter consisting of multimerized AP-1 site (AP-1 luciferase) to assess c-Jun transcriptional activity. The knock-down of *mbd3* induced an increase in reporter gene activity (Fig. 2a). Similar results were obtained using the urokinase-type plasminogen activator promoter (uPA-luciferase) (Supplementary Fig. 5a), a well known AP-1 target gene[14].

c-jun transcription is induced very quickly in response to JNK activation and c-Jun autoregulates its own promoter via two proximal AP-1 binding sites[15]. Chemical JNKi (Fig. 2b) or expression of *jnk1/2*-specific (also known as *MAPK8/9*) shRNAs, which resulted in substantial depletion of JNK protein (Supplementary Fig. 2c), decreased endogenous *c-jun* mRNA levels but *c-jun* transcriptional repression by JNK inhibition was impaired in the absence of Mbd3 (Fig. 2b and Supplementary Fig. 5b). Moreover, overexpression of Flag–Mbd3 decreased transcriptional activation of *c-jun* and, to a lesser extent, *cd44*, which is also regulated by c-Jun[4] (Supplementary Fig. 5d, e). Thus, depletion of Mbd3 results in derepression of AP-1 activity.

HCT116 colon cancer cells have constitutively high levels of activated JNK and phosphorylated c-Jun (Supplementary Fig. 2a) and chromatin immunoprecipitation (ChIP) for Mbd3 revealed minimal binding to a region covering the AP-1 sites of the *c-jun* promoter. However, JNKi and *jnk1/2* depletion resulted in a rapid and efficient recruitment of Mbd3 to the *c-jun* promoter (Fig. 2c and Supplementary Fig. 6a) and the *cd44* promoter (Supplementary Fig. 6b, c). Thus, the JNK signalling pathway regulates Mbd3 recruitment to c-Jun-dependent promoters.

ChIP analysis demonstrated that HDAC1 was also recruited to the *c-jun* promoter after JNKi and knock-down of Mbd3 prevented this recruitment (Fig. 2d). Unphosphorylatable c-Jun4A protein recruited Mbd3 more efficiently to the *c-jun* promoter compared to wild-type c-Jun (c-JunWT) (Supplementary Fig. 6d). c-Jun4A also recruited
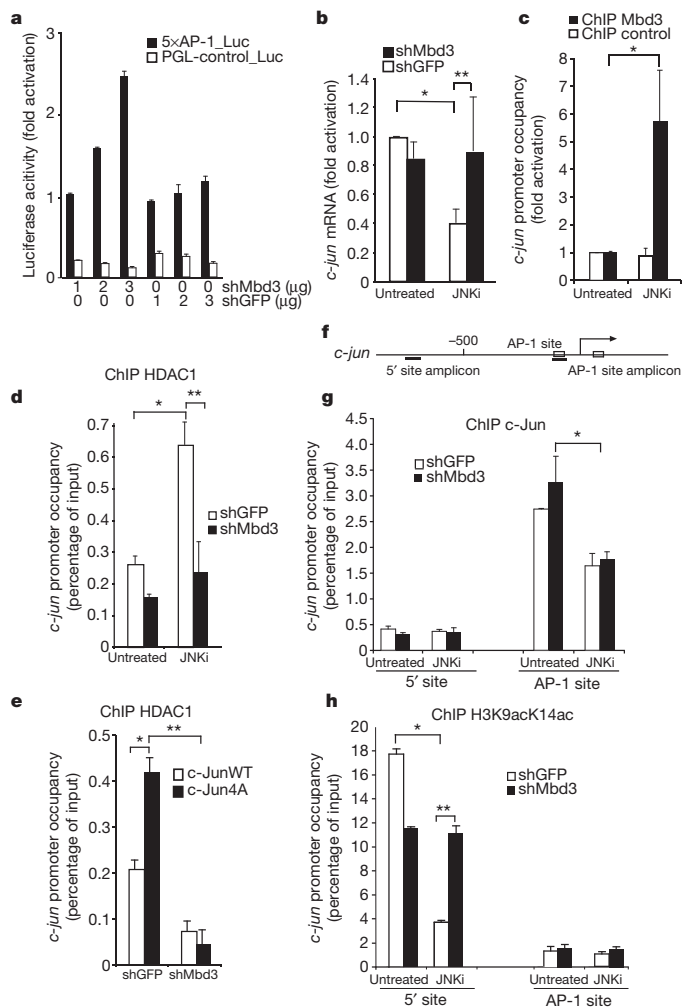
Figure 2 | Mbd3 represses *c-jun* transcription. a, Cells were transfected with a pSUPER vector expressing a shRNA specific for Mbd3 (shMbd3) or for GFP (shGFP). Activity of a 5×AP-1 luciferase reporter construct (5×AP-1_luc) or empty vector (PGL-control_Luc) relative to a thymidine kinase *Renilla* luciferase transfection control (tk-renilla) is shown ($n = 3$). **b**, mRNA levels of *c-jun* in the presence or absence of shMBD3 with JNKi or mock treatment ($n = 5$; $*P = 0.001$; $**P = 0.03$). **c**, ChIP for the AP-1 site in the *c-jun* promoter using Mbd3-specific or control antibody with or without JNKi ($n = 9$; $*P = 0.043$). **d**, HDAC1 ChIP for the AP-1 site in the *c-jun* promoter on cells expressing a shRNA specific for Mbd3 or for GFP ($*P = 0.004$; $**P = 0.0003$). **e**, HDAC1 ChIP for the *c-jun* AP-1 sites on cells expressing wild-type c-Jun (c-JunWT) or c-Jun4A ($*P = 0.02$; $**P = 0.01$). **f**, Schematic overview of the *c-jun* promoter. Underlined regions show the amplification region of each primer pair used. **g**, c-Jun ChIP for the 5′ and AP-1 sites of the *c-jun* locus on cells expressing a shRNA specific for Mbd3 or for GFP ($*P = 0.02$). **h**, ChIP for acetylated histone 3 K9, 14 (H3K9acK14ac) as in **g** ($*P = 0.0002$; $**P = 0.006$). In all experiments error bars show s.e.m. and Student's *t*-test was used for statistical analysis. HCT116 cells were used throughout. JNKi treatment was for 2 h in **b**, and for 30 min in **c**, **d**, **g** and **h**.

more HDAC1 to the *c-jun* promoter, and this recruitment was prevented when Mbd3 was silenced (Fig. 2e).

Histone acetylation has an important role in *c-jun* transcriptional regulation[16]. It has been shown previously that the regulation of histone acetylation at the *c-jun* promoter is complex. Increased histone acetylation in response to JNK activation was observed in a more distal region of the promoter (5′ site), whereas the region around the AP-1 sites remained unchanged[16] (Fig. 2f). In agreement with previous observations, we detected substantial histone H3 acetylation in the 5′ area of the promoter by ChIP, but no significant H3 acetylation around the AP-1 sites where c-Jun is localized (Fig. 2g, h). As expected, JNK inhibition induced a marked decrease in histone H3 acetylation,

but the regulation of H3 acetylation at the *c-jun* promoter was disrupted by Mbd3 depletion. H3 acetylation was slightly reduced in untreated cells, but in response to JNK inhibition no decrease in H3 acetylation was observed (Fig. 2h). Thus, Mbd3 contributes to the epigenetic regulation of the *c-jun* promoter by JNK signalling.

c-Jun has an important function in regulating intestinal epithelial homeostasis and intestinal cancer[3,4]. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis revealed an enrichment of *mbd3* mRNA in intestinal crypts (Supplementary Fig. 7a), and analysis of a knock-in mouse that expresses the reporter protein β-galactosidase under the control of the endogenous *mbd3* promoter (Supplementary Fig. 8a)[17] confirmed high *mdb3* expression in the crypt (Supplementary Fig. 7b). To address a potential role of Mbd3 in controlling c-Jun function in the intestine, we generated a mouse line in which exon 1 of Mbd3 was flanked with two *loxP* sites (Supplementary Fig. 8a). Exon 1 of Mbd3 encodes the N-terminal portion of the MBD and after Cre-mediated recombination *mbd3a* and *mbd3b*, the two transcripts

generated from the *mbd3* locus, are not produced (Supplementary Fig. 8b, c). We crossed the Mbd3 floxed mouse line with a Villin-Cre line[18] to specifically inactivate *mbd3* in the gut (*mbd3*$^{\Delta G/\Delta G}$). Deletion of the floxed *mbd3* allele was efficient (Supplementary Fig. 8d) and mRNA was isolated from crypts to investigate AP-1-dependent gene expression. Transcription of *c-jun* was increased in intestinal crypts lacking Mbd3, as was the transcription of several c-Jun target genes including *cd44* and *cyclind1* (also known as *Ccnd1*)(Fig. 3a). In addition, the expression of the stem cell marker *lgr5* was also increased in *mbd3*$^{\Delta G/\Delta G}$ intestine (Fig. 3a) whereas *bmi1* remained unchanged. *lgr5* marks crypt base columnar (CBC) cells, which function as intestinal stem cells[19]. The increase in *lgr5* mRNA was not due to an increase of CBC cells as their number was normal in the absence of Mbd3 (Fig. 3b). c-Jun is highly expressed in CBC cells[3,20], and *lgr5* expression is reduced in crypts lacking c-Jun[3]. We thus analysed whether *lgr5* might be directly regulated by c-Jun/Mbd3. *In silico* analysis revealed the presence of a consensus AP-1 site in intron 1 of the *lgr5* gene (Fig. 3c) and ChIP analysis
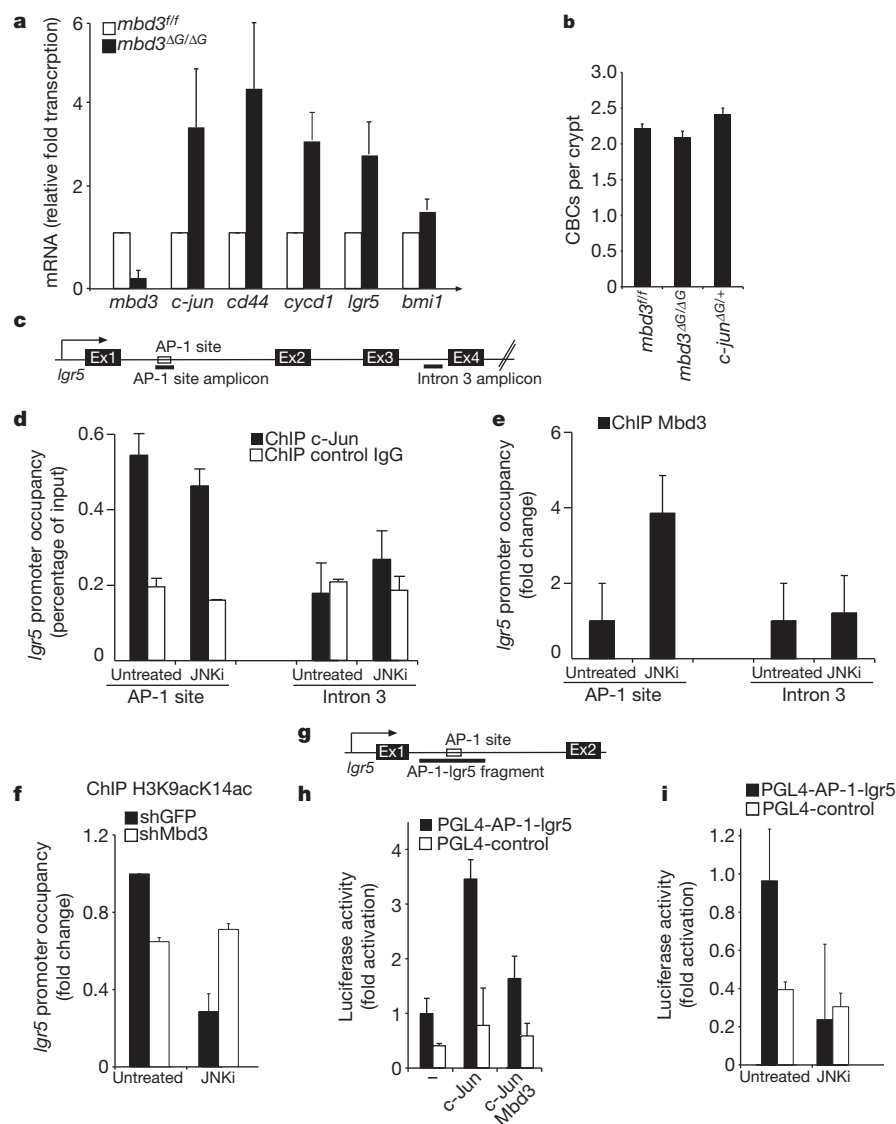

Figure 3 | *lgr5* is regulated by c-Jun-Mbd3/NuRD. a, qRT-PCR analysis of indicated genes on mRNA isolated from intestines of *mbd3*$^{f/f}$ and *mbd3*$^{\Delta G/\Delta G}$ mice ($n = 4$). Data were normalized to β-actin. **b**, Quantification of crypt base columnar cells (CBC) in *mbd3*$^{f/f}$ ($n = 4$), *mbd3*$^{\Delta G/\Delta G}$ ($n = 3$) and *mbd3*$^{\Delta G/\Delta G}$*c-jun*$^{\Delta G/+}$ ($n = 3$) mice. **c**, Schematic overview of the *lgr5* locus. Underlined regions show the amplification region of each primer pair used. **d**, ChIP for the *lgr5* AP-1 and intron 3 sites using c-Jun-specific or control antibody with or without JNKi. **e**, Mbd3 ChIP for the AP-1 and intron 3 sites in the *lgr5* locus.

f, Mbd3 ChIP for the *lgr5* AP-1 site with or without JNKi. **g**, Schematic overview of the fragment used for AP-1-lgr5 luciferase construct. **h**, **i**, Activity of an AP-1-lgr5 luciferase construct (PGL4-AP-1-lgr5) or empty vector (PGL4-control_Luc) relative to a tk-renilla luciferase transfection control in cells expressing c-Jun and Mbd3 (**h**) or with or without JNKi treatment (**i**). In all experiments (**c–f**, **h** and **i**) error bars show s.e.m. HCT116 cells were used and $n = 3$. JNKi treatment was for 30 min in **d–f** and for 5 h in **i**.

showed binding of c-Jun to this putative AP-1 site, but not to a region in intron 3 of the *lgr5* gene, used as a negative control (Fig. 3d). Mbd3 was also recruited to the same site in intron 1 of the *lgr5* gene, and JNKi increased Mbd3 recruitment (Fig. 3e). In addition, JNKi augmented CHD3 binding to the *lgr5* and *c-jun* AP-1 sites to a similar extent
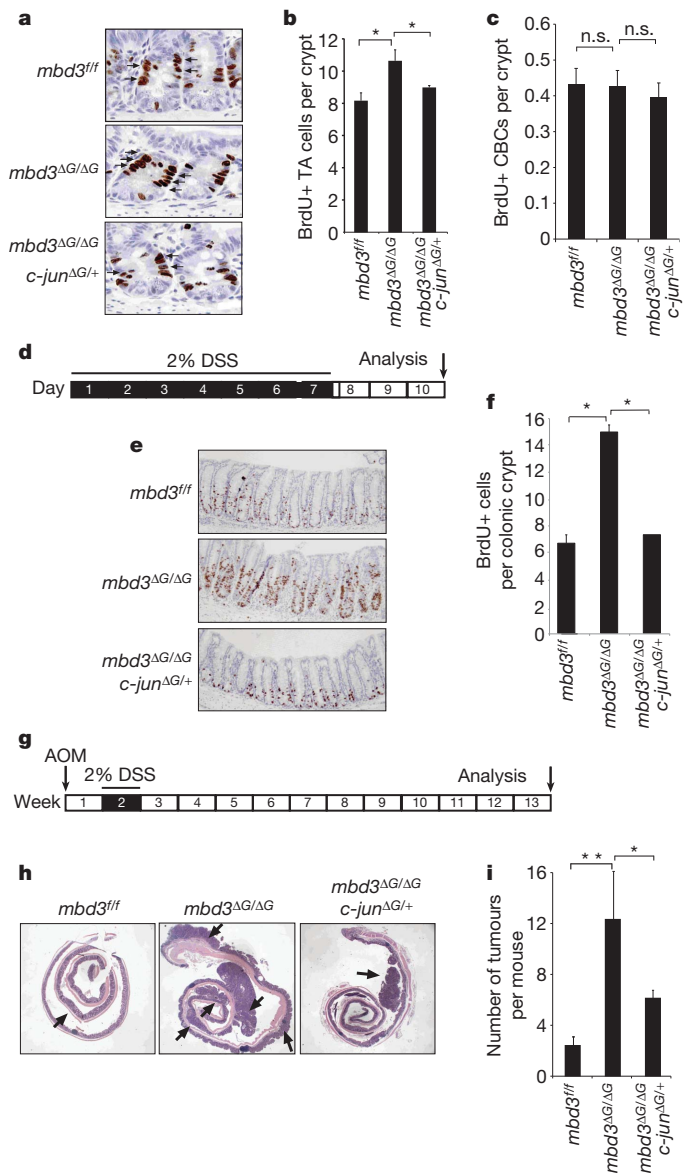

Figure 4 | Mbd3 antagonises c-Jun/AP-1 function *in vivo*.
a, Immunohistochemistry for BrdU+ cells on representative crypts of the indicated genotypes. Black arrows point to BrdU+ proliferative progenitors.
b, Quantification of BrdU-positive (BrdU+) transit amplifying (TA) cells in *mbd3*$^{f/f}$ ($n = 6$), *mbd3*$^{\Delta G/\Delta G}$ ($n = 5$) and *mbd3*$^{\Delta G/\Delta G}$*c-jun*$^{\Delta G/+}$ ($n = 4$) in intestinal crypts (error bars show s.e.m.; *$P = 0.001$; Student's *t*-test).
c, Quantification of BrdU+ CBCs in *mbd3*$^{f/f}$ ($n = 4$), *mbd3*$^{\Delta G/\Delta G}$ ($n = 3$) and *mbd3*$^{\Delta G/\Delta G}$*c-jun*$^{\Delta G/+}$ ($n = 3$) intestinal crypts. n.s., not significant.
d, Schematic representation of the experimental design for DSS-induced intestinal regeneration. **e**, Immunohistochemistry for BrdU+ cells on representative colonic crypts of mice with the indicated genotypes.
f, Quantification of BrdU+ cells in *mbd3*$^{f/f}$ ($n = 10$), *mbd3*$^{\Delta G/\Delta G}$ ($n = 6$) and *mbd3*$^{\Delta G/\Delta G}$*c-jun*$^{\Delta G/+}$ ($n = 2$) mice colonic crypts; *$P = 0.00012$. **g**, Schematic representation of the experimental design for colitis-induced tumorigenesis.
h, Haematoxylin and eosin staining on representative colons of mice with the indicated genotypes. **i**, Quantification of number of AOM/DSS-induced tumours in *mbd3*$^{f/f}$ ($n = 5$), *mbd3*$^{\Delta G/\Delta G}$ ($n = 4$) and *mbd3*$^{\Delta G/\Delta G}$*c-jun*$^{\Delta G/+}$ ($n = 2$) mice (**$P = 0.01$, *$P = 0.016$). In all experiments error bars show s.e.m. and Student's *t*-test was used for statistical analysis.

(Supplementary Fig. 6e, f). JNK inhibition reduced acetylation of the *lgr5* AP-1 site but the Mbd3 depletion impaired JNKi-induced deacetylation (Fig. 3f). When inserted into a luciferase reporter construct, the *lgr5* fragment containing the AP-1 site (pGL4-Ap-1-lgr5) (Fig. 3g) was activated by c-Jun overexpression whereas coexpression with Mbd3 resulted in repression (Fig. 3h). JNKi reduced pGL4-Ap-1-lgr5 reporter activity to background levels (Fig. 3i). These data suggest that Mbd3/c-Jun regulates *lgr5* transcription.

The only described biological function of Mbd3 is in controlling proliferation and differentiation of embryonic stem cells[21,22], so we investigated whether intestinal cell lineage specification was impaired in the gut of *mbd3*$^{\Delta G/\Delta G}$ mice. Immunohistochemical stainings for Alcian blue/periodic acid Schiff (AB/PAS) (which detects goblet cells), chromogranin (for enteroendocrine cells) or lysozyme (for Paneth cells) failed to detect a defect in cell fate specification in the intestine of *mbd3*$^{\Delta G/\Delta G}$ mice (Supplementary Fig. 9a–c). Interestingly, *mbd3*$^{\Delta G/\Delta G}$ mice showed a significant increase in bromodeoxyuridine (BrdU)-positive transient amplifying cells in the crypts compared to controls (Fig. 4a, b), but we did not detect any increase in CBC cell proliferation (Fig. 4c). To test whether increased c-Jun activity was responsible for the augmented intestinal progenitor proliferation in *mbd3*$^{\Delta G/\Delta G}$ mice, one *c-jun* allele was deleted in a *mbd3*$^{\Delta G/\Delta G}$ background, generating *mbd3*$^{\Delta G/\Delta G}$; *c-jun*$^{\Delta G/+}$ mice. The hyperproliferation phenotype in the absence of Mbd3 was rescued in *mbd3*$^{\Delta G/\Delta G}$; *c-jun*$^{\Delta G/+}$ mice (Fig. 4a, b). We next investigated the impact of *mbd3* inactivation under pathological conditions. Intestinal colitis and regeneration was induced by dextran sodium sulphate (DSS) administration (Fig. 4d). *mbd3*$^{\Delta G/\Delta G}$ mice showed a large increase in proliferating cells in the colon after DSS treatment compared to control animals. The hyperproliferation phenotype was completely reversed by deletion of a single *c-jun* allele (Fig. 4e, f). Moreover, to investigate a potential role of Mbd3 in intestinal tumorigenesis, we applied the azoxymethane (AOM)/DSS protocol to trigger colitis-induced tumorigenesis (Fig. 4g)[3]. Mbd3 deletion resulted in a striking increase in tumour number (Fig. 4h, i) as well as an increase in tumour size (Supplementary Fig. 9d) that was significantly rescued by concomitant deletion of one *c-jun* allele (Fig. 4h, i and Supplementary Fig. 9d). Thus, in the intestine Mbd3 controls cellular proliferation under physiological and pathological conditions by antagonising c-Jun activity.

The direct recruitment of the NuRD complex by Mbd3 interaction with unphosphorylated c-Jun suggests that in the absence of JNK pathway activation, c-Jun actively represses AP-1 target gene expression, providing a molecular explanation for previous observations that c-Jun phosphorylation mediates dissociation of an inhibitory complex associated with histone deacetylation[23,24]. The phosphorylation-dependent interaction of Mbd3/c-Jun is in agreement with a three-step mechanism of c-Jun activation by JNK. Upon JNK activation, NuRD dissociates from c-Jun, resulting in derepression of target gene transcription. Subsequently, a coactivator protein would bind to the c-Jun TAD and further augment gene expression. It is worth noting that to explain stimulation of c-Jun activity by N-terminal phosphorylation, coactivator binding may, but is not required to be phosphorylation-dependent[4,25]. Lastly, upon termination of the JNK signal, Mbd3/NuRD would re-bind to unphosphorylated c-Jun resulting in cessation of target gene expression. Thus, the TAD of c-Jun functions as a repressor domain that is negatively regulated by phosphorylation.

METHODS SUMMARY

Cell culture. HCT116 and HEK293T cells were grown in DMEM and 10% FBS.
IP, pull-down and western blot analysis. Cells were lysed in the appropriate buffer for immunoprecipitation, GST pull-down or peptide pull-down assays. Immunoblots were carried out as previously described[4].
Reporter gene assay. Firefly and *Renilla* luciferase reporter assays were performed and measured using the Dual-Luciferase Reporter Assay System (Promega).
Chromatin immunoprecipitation (ChIP). ChIP analysis was performed as described previously[26]. qRT-PCR was accomplished with SYBR Green incorporation.

Mouse lines. Mbd3$^{\Delta G/\Delta G}$ mice resulted from crossing Mbd3 floxed line[21] with villin cre line[18].

AOM/DSS model of colon carcinoma and DSS model of colon inflammation. The AOM/DSS model[27] and the DSS protocol used[28] have been previously described. For tissue collection, mice were sacrificed and the intestines were frozen in liquid nitrogen. For qRT–PCR analysis, total mRNA was isolated from dissected intestines, or from villus and crypt fractions obtained as described before[19].

Immunohistochemical staining. For haematoxylin & eosin staining and immunohistochemistry, the intestines were fixed overnight in 10% neutral buffered formalin, transferred into 70% ethanol, rolled, processed and embedded into paraffin. Sections were cut at 4 μm.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Davis, R. J. Signal transduction by the JNK group of MAP kinases. *Cell* **103,** 239–252 (2000).
2. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nature Rev. Cancer* **3,** 859–868 (2003).
3. Sancho, R. *et al.* JNK signalling modulates intestinal homeostasis and tumourigenesis in mice. *EMBO J.* **28,** 1843–1854 (2009).
4. Nateri, A. S., Spencer-Dene, B. & Behrens, A. Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. *Nature* **437,** 281–285 (2005).
5. Pulverer, B. J., Kyriakis, J. M., Avruch, J., Nikolakaki, E. & Woodgett, J. R. Phosphorylation of c-jun mediated by MAP kinases. *Nature* **353,** 670–674 (1991).
6. Behrens, A., Sibilia, M. & Wagner, E. F. Amino-terminal phosphorylation of c-jun regulates stress-induced apoptosis and cellular proliferation. *Nature Genet.* **21,** 326–329 (1999).
7. Nateri, A. S., Riera-Sans, L., Da Costa, C. & Behrens, A. The ubiquitin ligase SCFFbw7 antagonizes apoptotic JNK signaling. *Science* **303,** 1374–1378 (2004).
8. Le Guezennec, X. *et al.* MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26,** 843–851 (2006).
9. Denslow, S. A. & Wade, P. A. The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26,** 5433–5438 (2007).
10. Ahringer, J. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet.* **16,** 351–356 (2000).
11. Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. Mi-2/NuRD: multiple complexes for many purposes. *Biochim. Biophys. Acta* **1677,** 52–57 (2004).
12. Saito, M. & Ishikawa, F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* **277,** 35434–35439 (2002).
13. Hendrich, B. & Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* **19,** 269–277 (2003).
14. D'Orazio, D. *et al.* Cooperation of two PEA3/AP1 sites in uPA gene induction by TPA and FGF-2. *Gene* **201,** 179–187 (1997).
15. Angel, P., Hattori, K., Smeal, T. & Karin, M. The *jun* proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell* **55,** 875–885 (1988).
16. Clayton, A. L., Rose, S., Barratt, M. J. & Mahadevan, L. C. Phosphoacetylation of histone H3 on c-*fos*- and c-*jun*-associated nucleosomes upon gene activation. *EMBO J.* **19,** 3714–3726 (2000).
17. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* **15,** 710–723 (2001).
18. El Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39,** 186–193 (2004).
19. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
20. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136,** 903–912 (2009).
21. Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8,** 285–292 (2006).
22. Kaji, K., Nichols, J. & Hendrich, B. Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* **134,** 1123–1132 (2007).
23. Baichwal, V. R., Park, A. & Tjian, R. The cell-type-specific activator region of c-Jun juxtaposes constitutive and negatively regulated domains. *Genes Dev.* **6,** 1493–1502 (1992).
24. Weiss, C. *et al.* JNK phosphorylation relieves HDAC3-dependent suppression of the transcriptional activity of c-Jun. *EMBO J.* **22,** 3686–3695 (2003).
25. Arias, J. *et al.* Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature* **370,** 226–229 (1994).
26. Nelson, J. D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature Protocols* **1,** 179–185 (2006).
27. Neufert, C., Becker, C. & Neurath, M. F. An inducible mouse model of colon carcinogenesis for the analysis of sporadic and inflammation-driven tumor progression. *Nature Protocols* **2,** 1998–2004 (2007).
28. Floer, M. *et al.* Enoxaparin improves the course of dextran sodium sulfate-induced colitis in syndecan-1-deficient mice. *Am. J. Pathol.* **176,** 146–157 (2010).

Author Contributions C.A. designed and performed most of the experiments, analysed data, and co-wrote the paper. K.N. identified Mbd3 as binding specifically to unphosphorylated c-Jun using the yeast three-hybrid screen, generated reagents and provided Fig. 1a, c and Supplementary Fig. 3. R.S. helped with all *in vivo* experiments and provided Fig. 3a and Supplementary Fig. 7a. A.C. provided Fig. 1f–i. B.H. generated the *mbd3* floxed mouse line. A.B. supervised all aspects of this work and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to axel.behrens@cancer.org.uk

METHODS

Cell culture and transfection. HCT116 and HEK293T cells were cultured in Dulbeccos's modified Eagle medium (DMEM) and 10% FBS. Cells were plated at subconfluence and transfected with lipofectamine reagent (Invitrogen). HCT116 cells were treated when indicated with 50 μM JNKi (SP600125, Calbiochem) or 25 ng ml^{-1} anisomycin (Sigma).

Immunoprecipitation. Cells were lysed for 30 min at 4 °C in 500 μl of immuno-precipitation buffer containing 0.5% Triton X-100, 1 mM EDTA, 100 μM Na-orthovanadate, 0.25 mM PMSF (phenylmethylsulphonyl fluoride), and protease inhibitor mixture (Sigma) in PBS. After centrifugation, supernatants were incubated for 3 h at 4 °C with 1 μg of the anti-Flag (Sigma) antibody coupled to protein A-Sepharose beads.

GST pull-down. GST fusion proteins were purified from the *Escherichia coli* BL-21 strain in a buffer containing 20 mM Tris-HCl, pH 7.4, 1 M NaCl, 0.2 mM EDTA, 1 mM dithiothreitol (DTT), 1 mg ml^{-1} lysozyme, 1 mM PMSF, and protease inhibitor complex (Sigma) and bound to glutathione-Sepharose (Amersham Biosciences). GST fusion proteins were purified and incubated with 400 μg of cell lysates for 2 h at 4 °C in RIPA buffer and extensively washed. Pulled down proteins were analysed by western blot.

Peptide pull-down. Biotinylated peptide (1 mg) was incubated with Dynabeads M280 streptavidin in buffer A (5 mM Tris-Cl pH 7.5, 0.5 mM EDTA, 1 M NaCl) at 4 °C overnight. Peptide bound beads were washed and resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.2% NP 40, 1 mM DTT). 293T cells overexpressing c-Jun were lysed in lysis buffer and then incubated with different Dynabead-bound peptides for 2 h at 4 °C. The beads were washed three times with lysis buffer containing 400 mM NaCl. Pulled-down proteins were analysed by western blot.

Western blot analysis. HCT116 cells lysates were homogenized in RIPA lysis buffer supplemented with protease inhibitor (Sigma). Immunoblots were carried out as described previously[4]. The gels were transferred to nitrocellulose membranes, and the membranes were immunoblotted with various antibodies as indicated: anti-c-Jun (H79, Santa Cruz sc1694 and BD Transduction Laboratories 610326), anti-Myc (9E10 clone), anti-Flag–HRP (horseradish peroxidase-conjugated; Sigma), anti-c-Jun-ser63 and anti-β-Actin (Sigma), anti-Mbd3 (c-18, Santa Cruz sc-9402), anti-JNK1/2 (NEB cell signalling 9252), anti-CHD3 (BD Transduction Laboratories 611847) and anti-MTA3 (Santa Cruz sc- 48799). HCT116 cells were treated where indicated with ultraviolet at 40 J cm^{-2}.

Reporter gene assay. HCT116 cells were transfected with the indicated plasmids with lipofectamine reagent (Invitrogen). Transient transfections of the indicated plasmids and controls of firefly and *Renilla* luciferase reporters were performed and measured using the Dual-Luciferase Reporter Assay System (Promega), 36 h post-transfection. Data are expressed as fold induction after being normalized using tk-renilla luciferase (mean ± s.e.m.; $n = 3$).

Chromatin immunoprecipitation (ChIP). ChIP analysis was performed as described previously[26]. Cells were treated as indicated before collection, with 50 μM JNKi (SP600125, Calbiochem), 25 ng ml^{-1} anisomycin (Sigma) or transfected with the indicated plasmids with lipofectamine reagent (Invitrogen). Immunoprecipitations were carried out with the indicated antibodies, anti-Mbd3 (c-18) (Santa Cruz sc9402), anti-HDAC1 (H51) (Santa Cruz sc7872), anti-c-Jun (H79) (Santa Cruz sc1694) or anti-acetyl K9, K14 histone 3 (Upstate 06-599) and collected by agarose beads (Santa-Cruz). The oligonucleotide sequences used to amplify the DNA fragments by qRT-PCR are AP1 c-Jun F, 5′-GCGACGCGAGCCAATG-3′; AP1 c-Jun R, 5′-AGCCCGAGCTCAACAC TTATCT-3′; 5′ c-Jun F, 5′-TTATACTCCTCCTCCCCCAGCTA-3′; 5′ c-Jun R, 5′-CCTCCAGCCTCCTGAAACATC-3′; lgr5-AP-1 F, 5′-TGGGAATCA GTGCGGTTCA-3′; lgr5-AP-1 R, 5′-CAGGGAAATGGGCAAACAAG-3′; lgr5-intron3 F, 5′-TCTGCCTCAGGCTTACATGGA-3′; lgr5-intron3 R, 5′-CA CAAGAATTCTGCAGCACATTT-3′; F-CD44 AP-1, 5′-AGTGCAGTGGCACG ATCTTG-3′; R-CD44 AP-1, 5′-CGGGAGAATCGCTTGAACAC-3′.

Quantitative real-time PCR was performed with SYBR Green incorporation (Platinum Quantitative PCR SuperMix-UDG w/ROX, Invitrogen) using an ABI7900HT (Applied Bioscience) and the data were analysed using the SDS 2.3 software.

RNA extraction and qRT-PCR analysis. Total mRNA was isolated using RNeasy Mini-kit according to the manufacturer's instructions (Qiagen). cDNA was synthesized using Invitrogen Superscript reagents according to the manufacturer's instructions. The oligonucleotide sequences used to amplify the DNA fragments by qRT-PCR are F-h-c-Jun, 5′-TCGACATGGAGTCCCAGGA-3′; R-h-c-Jun, 5′-GGCGATTCTCTCCAGCTTCC-3′; F-h-actin, 5′-GGATGCAGAAGGAG ATCACTG-3′; R-h-actin, 5′-CGATCCACACGGAGTACTTG-3′; F-hCD44, 5′-CTCCTGGCACTGGCTCTGA-3′; R-hCD44, 5′-CTGCCCACACCTTCT CCTACTATT-3′; F-mMbd3, 5′-CACCGCGCTGCCTGTAC-3′; R-mMbd3, 5′-TGGTCACCGGTTGCTTGA-3′; F-mc-Jun, 5′-TGAAAGCTGTGTCCCC TGTC-3′; R-mc-Jun, 5′-ATCACAGCACATGCCACTTC-3′; F-mβ-actin, 5′-ATGCTCCCCGGGCTGTAT-3′; R-mβ-actin, 5′-CATAGGGAGTCCTTCTG ACCCATTC-3′; F-mCD44, 5′-CTCCTGGCACTGGCTCTGA-3′; R-mCD44, 5′-CTGCCCACACCTTCTCCTACTATT-3′; F-mCCDN1, 5′-GTGCGTGCA GAAGGAGATTGT-3′; R-mCCDN1, 5′-CTCACAGACCTCCAGCATCCA-3′; F-mLgr5, 5′-CGGAGGAAGCGCTACAGAAT-3′; R-mLgr5, 5′-CTGGGT GGCACGTAGCTGAT-3′; F-mBmiI, 5′-GGGCTTTTCAAAAATGAGATGAA-3′; R-mBmiI, 5′-GAGCCATTGGCAGCATCAG-3′.

Quantitative real-time PCR was accomplished with SYBR Green incorporation (Platinum Quantitative PCR SuperMix-UDG w/ROX, Invitrogen) using an ABI7900HT (Applied Bioscience), and the data were analysed using the SDS 2.3 software. Results were normalized to those obtained with β-actin.

RNAi constructs. siRNA sequences directed towards the coding region of the indicated genes were designed using the Dharmacon siDESIGN centre. Annealed hairpin oligonucleotides were cloned into pSuper (OligoEngine).

shGFP, sense strand 5′-GAACGGCATCAAGGTGAAC-3′; shJNK1/2, sense strand 5′-AAAGAATGTCCTACCTTCTT-3′; shMbd3-a, sense strand 5′-AGA CGGCGTCCATCTTCAA-3′; shMbd3-b, sense strand 5′-TATGGCTCAAC ACCACGCA-3′; shMbd3-c, sense strand 5′-GGACATCAGGAAGCAGGAA-3′.

Mouse lines. Embryonic stem cells in which one *mbd3* allele is deleted and the other contains *loxP* sites flanking exon 1 have been described[21]. Upon transfection with Cre recombinase, the resulting *mbd3 Ex1$^{Δ/−}$* embryonic stem cells lack Mbd3a and Mbd3b. *mbd3$^{Flox/−}$* embryonic stem cells were used for blastocyst injection to generate a mouse line homozygous for the *mbd3$^{Ex1-Flox}$* allele. These mice were viable and fertile. The Villin-Cre line has been described previously[18]. For tissue collection, mice were killed by cervical dislocation and the intestines were removed and flushed extensively with cold PBS and frozen in liquid nitrogen. For quantitative real-time PCR (qRT–PCR) analysis, total mRNA was isolated from dissected intestines, or from villus and crypt fractions obtained as described before[19]. Results were normalized to those obtained with β-actin and results are presented as fold induction over control mice.

AOM/DSS model of colon carcinoma. The AOM/DSS model used has been described previously[27]. Mice were injected intraperitoneally (i.p.) with 10 mg kg^{-1} body weight of AOM (Sigma) dissolved in physiological saline. Seven days later, 2% DSS was given in the drinking water over 7 days, followed by normal water until the end of the experiment. Body weight was measured every week and the animals were killed 13 weeks after AOM injection for histological analysis. The number of tumours, incidence of tumours and tumour area were determined. For biochemical characterization dissected tumours were snap-frozen in liquid nitrogen and used for qRT–PCR analysis.

DSS model of colon inflammation. The DSS model used has been described previously[28]. Mice were administered orally with 2% DSS in drinking water starting on day 1 and continued until day 6. On day 7, the water was changed to normal drinking water. Body weight was measured every day, and the animals were killed at day 10 for histological analysis.

Immunohistochemical staining. Mice were injected i.p. with 100 mg kg^{-1} BrdU (Sigma) 1.5 h before killing. Mice were euthanized by cervical dislocation, the small intestines dissected out into ice-cold PBS and the faecal contents flushed out with ice-cold PBS. The intestines were cut longitudinally into pieces of similar size, opened out and fixed overnight in 10% neutral buffered formalin, briefly washed with PBS and transferred into 70% ethanol, rolled, processed and embedded into paraffin. Sections were cut at 4 μm for haematoxylin and eosin staining. For immunohistochemistry, antibodies against BrdU (BD Biosciences) were used. To quantify the BrdU-positive cells per crypt, 100 full crypts were scored. For AB/PAS staining, immunohistochemistry and immunofluorescence antibodies against chromogranin-A (Abcam) and β-galactosidase (Acris Antibodies) were used. Anti-lysozyme (DAKO) was used for immunofluorescence as described before[4]. To quantify the AB/PAS$^+$ or chromogranin$^+$ cells per villus, 100 full villi were scored.

Statistics. Statistical evaluation was performed using the Student's unpaired t-test. Data are presented as mean ± s.e.m. and $P \leq 0.05$ was considered statistically significant.

LETTER

Ependymal cells of chordate larvae are stem–like cells that form the adult nervous system

Takeo Horie[1], Ryoko Shinki[1], Yosuke Ogura[1], Takehiro G. Kusakabe[2], Nori Satoh[3] & Yasunori Sasakura[1]

In ascidian tunicates, the metamorphic transition from larva to adult is accompanied by dynamic changes in the body plan. For instance, the central nervous system (CNS) is subjected to extensive rearrangement because its regulating larval organs are lost and new adult organs are created[1]. To understand how the adult CNS is reconstructed, we traced the fate of larval CNS cells during ascidian metamorphosis by using transgenic animals and imaging technologies with photoconvertible fluorescent proteins[2]. Here we show that most parts of the ascidian larval CNS, except for the tail nerve cord, are maintained during metamorphosis and recruited to form the adult CNS. We also show that most of the larval neurons disappear and only a subset of cholinergic motor neurons and glutamatergic neurons are retained. Finally, we demonstrate that ependymal cells of the larval CNS contribute to the construction of the adult CNS and that some differentiate into neurons in the adult CNS. An unexpected role of ependymal cells highlighted by this study is that they serve as neural stem-like cells to reconstruct the adult nervous network during chordate metamorphosis. Consequently, the plasticity of non-neuronal ependymal cells and neuronal cells in chordates should be re-examined by future studies[3,4].

Among chordates, urochordate ascidians perform dramatic metamorphosis, which converts swimming larvae to sessile adults[5]. The larva of ascidians contains a well-organized tripartite brain[6]. The traditional view, based on morphological studies, is that most of the larval brain is destroyed during metamorphosis. The adult CNS is then formed from the neurohypophysial duct, which is located in the anterior-most part of the larval brain connecting to the stomodeum[7,8]. However, more recent work using yellow fluorescent protein has suggested that, for *Ciona* at least, a part of the larval CNS, notably the neck region, is an important source of adult neurons[9]. Our intent here is to extend this earlier study to address more definitively how much of the larval CNS is lost and how much of it contributes to form the adult CNS during ascidian metamorphosis.

To address this, we used Kaede fluorescent protein because Kaede fluorescence can be irreversibly converted from green to red by irradiation with ultraviolet light[2] (Fig. 1a). Consequently, we can label cells in the larval CNS with red fluorescence and trace them during metamorphosis (Fig. 1). The newly formed neural cells can be distinguished from the larval neural cells because of their exclusive transmission of green fluorescence. We have established a transgenic line of *Ciona intestinalis* that expresses Kaede in the entire larval nervous system with the *cis* regulatory element of the *Ci-β2tubulin* gene[10] (Fig. 1b). Ultraviolet irradiation was applied to larvae of the transgenic line (Fig. 1c), which were observed during and after metamorphosis. We found that red-labelled cells derived from the larval CNS were located in the entire region of the cerebral ganglion and ciliated funnel, two major parts of the adult CNS (Fig. 1d). This indicates that cells in the adult CNS are inherited from the larval CNS. Most red-labelled cells also expressed green fluorescence (Fig. 1d), suggesting their continuous expression of Kaede from the *β2tubulin* regulatory element. In contrast,

cells around pigment cells, which were located in the middle part of the cerebral ganglion, showed strong red fluorescence but only background levels of green fluorescence (Fig. 1d, asterisk). This suggests that these cells terminated their expression of the Kaede protein from the *β2tubulin* regulatory element during metamorphosis. Support for this notion
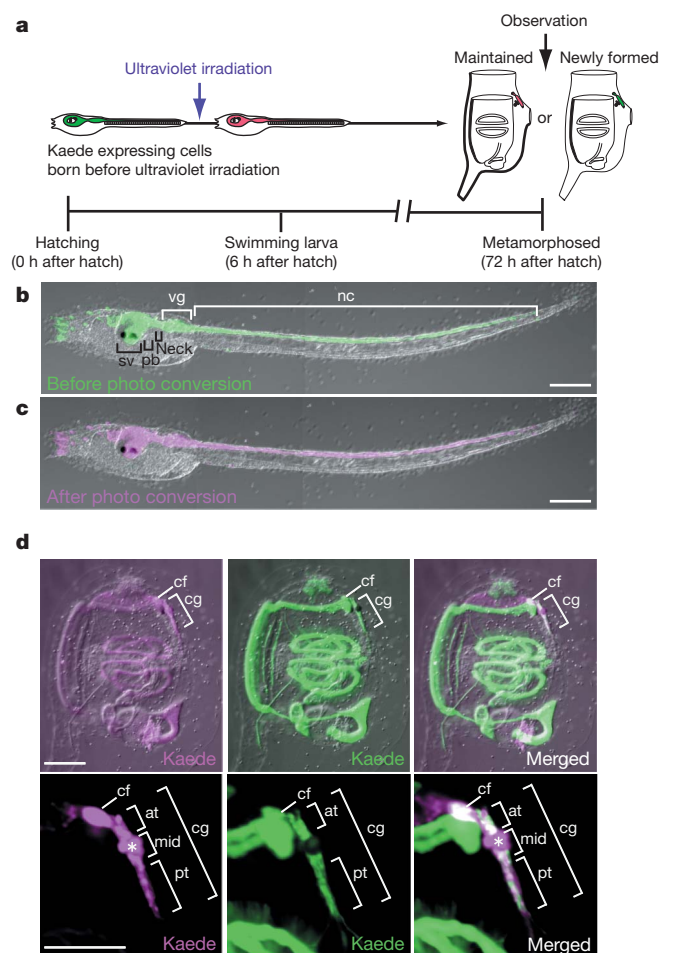

Figure 1 | Tracing of cells in the larval CNS. a, The tracing experiment. **b**, Kaede expression in the larval CNS of the *β2tubulin*-Kaede line; sv, sensory vesicle; pb, posterior brain; vg, visceral ganglion; nc, nerve cord. **c**, A larva after ultraviolet irradiation. Red fluorescence is shown in magenta. **d**, Distribution of larval neural cells in the adult CNS. Upper, photographs of a whole juvenile; lower, magnified images of the adult CNS. Left, Kaede red fluorescence; middle, Kaede green fluorescence; right, merged images. Cells with red fluorescence are distributed throughout the cerebral ganglion (cg) and ciliated funnel (cf). Asterisk indicates cells around pigment cells. The cerebral ganglion is divided into anterior (at), middle (mid) and posterior (pt) parts. Kaede green fluorescence is also observed in the mesodermal and endodermal tissues. Scale bars, 100 μm.

[1]Shimoda Marine Research Center, University of Tsukuba, Shimoda, Shizuoka 415-0025, Japan. [2]Department of Biology, Konan University, Kobe, Hyogo 658-8501, Japan. [3]Marine Genomics Unit, Okinawa Institute of Science and Technology Promotion Corporation, Onna, Okinawa 904-0412, Japan.

comes from a previous report that cells around the pigment cells are subjected to programmed cell death during metamorphosis[11]. Therefore the red cells around the pigment cells are highly likely to be dead larval cells, so we excluded these from subsequent analyses. The cerebral ganglion contained not only red-labelled cells but also cells with only green fluorescence that were newly born after the larval stage (Fig. 1d). However, such new cells were evidently fewer than larval neural cells, suggesting that most of the cells composing the adult CNS were derived from the larval CNS.

The larval CNS is largely divided into five parts: the sensory vesicle (the anterior part of the brain), posterior brain, neck, visceral ganglion and nerve cord (Fig. 1b). Using ultraviolet irradiation of specific parts of the larval CNS in the *β2tubulin*-Kaede line, we examined which of these five parts was maintained to participate in adult CNS formation. When cells in the sensory vesicle and posterior brain were traced, red-labelled cells were present in the entire cerebral ganglion and ciliated funnel (Supplementary Fig. 1a). This suggests that cells in the sensory vesicle/posterior brain are the major source of cells composing the adult CNS. When non-neuronal cells in the posterior brain and neck were traced separately from those of the sensory vesicle (by expressing Kaede with the *cis* element of *CiPhox2* (ref. 9); Supplementary Fig. 1b), the cells remained to form the relatively posterior part of the cerebral ganglion (Supplementary Fig. 1b); this suggests the conservation of the anterior–posterior axis between the larval and adult CNS. When cells in the visceral ganglion were traced, a few red-labelled cells were present in the middle and posterior parts of the cerebral ganglion (Supplementary Fig. 1c). In contrast, cells in the tail nerve cord disappeared completely after metamorphosis, indicating no contribution of these cells to adult CNS formation (Supplementary Fig. 1d).

The larval CNS of *Ciona* is composed of approximately 350 cells, including five sensory receptor cells, approximately 100 neurons and approximately 245 ependymal cells[12,13]. Because most of the larval neurons are cholinergic, glutamatergic or GABAergic/glycinergic[14–16], we determined which type of neuron contributes to the adult CNS. To trace the location of larval cholinergic neurons, we used our previously established transgenic line that expresses Kaede in cholinergic neurons with the upstream regulatory region of *Ci-VACHT*, a specific marker gene of the neurons[17,18]. Larvae of the *VACHT*-Kaede line expressed Kaede in the posterior brain and the motor neurons in the visceral ganglion (Fig. 2a). When larval cholinergic neurons were traced during metamorphosis, red-labelled, larval cholinergic neurons were found in the posterior part of the cerebral ganglion (Fig. 2b). Red-coloured cholinergic neurons extended long axons posteriorly, suggesting that they are functional neurons in the adult CNS. We then examined whether the cholinergic neurons in the posterior brain or the visceral ganglion were linked to the adult neurons. We found that only the cholinergic motor neurons in the visceral ganglion remained in the posterior part of the cerebral ganglion (Supplementary Fig. 2).

We then traced glutamatergic neurons during metamorphosis. Again, we used a previously established transgenic line that expresses Kaede in glutamatergic neurons by the promoter of *Ci-VGLUT*, a gene expressed specifically in glutamatergic neurons[14] (Supplementary Fig. 3a). When larval glutamatergic neurons were traced, the anterior-most part of the ciliated funnel showed both red and green fluorescence, suggesting that this part contains larval glutamatergic neurons (Supplementary Fig. 3b). Therefore most of the larval glutamatergic neurons were lost during metamorphosis, although a few of them remained and formed the ciliated funnel.

We found a different result when tracing GABAergic/glycinergic neurons. GABAergic/glycinergic neurons were labelled with the upstream regulatory region of their marker gene *Ci-VGAT*[18] (Supplementary Fig. 3c). When the larval GABAergic/glycinergic neurons were traced during metamorphosis, a conspicuous red fluorescence was detected in dead cells around the pigmented cells (Supplementary Fig. 3d). This suggests that larval GABAergic/glycinergic neurons disappeared during metamorphosis.
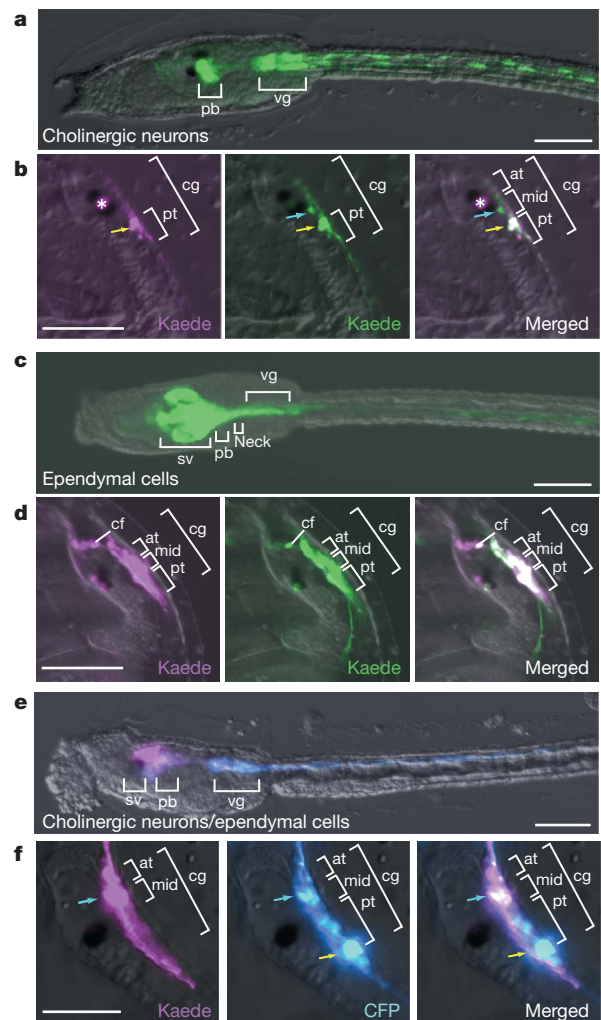

Figure 2 | Tracing of larval cholinergic neurons and ependymal cells, and differentiation of larval ependymal cells to adult cholinergic neurons. a, Kaede expression in the posterior brain (pb) and visceral ganglion (vg) of a larva of the *VACHT*-Kaede line. **b**, Tracing of cholinergic neurons. Larval cholinergic neurons are located in the posterior (pt) part of the cerebral ganglion (cg) (yellow arrow). The asterisk indicates dead cells around the pigmented cells. Newly born neural cells are present in the middle part (blue arrow). **c**, A larva into which the *Ci-opsin3* promoter and Kaede fusion construct was introduced by electroporation. Kaede is expressed in the sensory vesicle (sv), posterior brain (pb), neck and visceral ganglion (vg). **d**, Tracing of ependymal cells. Red-coloured cells are present in the entire region of the cerebral ganglion (cg) and ciliated funnel (cf). **e**, Expression of CFP and photoconverted Kaede in the larval cholinergic neurons and ependymal cells, respectively. **f**, Kaede and CFP fluorescence in the adult cerebral ganglion after metamorphosis. Cholinergic neurons in the anterior to middle part of the cerebral ganglion (blue arrow) showed Kaede red fluorescence, suggesting that they were derived from larval ependymal cells. Cholinergic neurons in the posterior part (yellow arrow) did not show Kaede red fluorescence. Scale bars, 100 μm.

We also followed the outcome of glial cells. Ependymal cells, which line the cavity of the sensory vesicle and constitute the entire nerve cord[12,19], are the only glial cell type found in the ascidian CNS. The *Ciona* larval brain has approximately 70% of these cells[13]. Kaede was expressed in ependymal cells through the upstream regions of *Ci-opsin3* and *Ci-CRALBP*, marker genes of ependymal cells[20] (Fig. 2c). The trace of larval ependymal cells revealed the distribution of red-labelled cells throughout the cerebral ganglion and ciliated funnel (Fig. 2d). When ependymal cells in the visceral ganglion were specifically traced (by DiI labelling), these cells remained after metamorphosis in the posterior part of the adult cerebral ganglion (Supplementary Fig. 4).

This finding supports the notion that there is conservation of the anterior–posterior axis between the larval and adult CNS. Next, to determine whether adult neurons are produced from larval ependymal cells, their fate after metamorphosis was traced by the co-expression of Kaede in the larval ependymal cells and a cyan fluorescent protein (CFP) with neuron subtype-specific promoters (Supplementary Fig. 5a). When cholinergic neurons were examined, it was found that the CFP-positive cholinergic neurons located in the anterior part of the cerebral ganglion showed red Kaede fluorescence, the marker of larval ependymal cells (Fig. 2e, f, blue arrow). This indicates that they were derived from larval ependymal cells. Posterior cholinergic neurons were negative for Kaede fluorescence (Fig. 2f, yellow arrow), confirming that they are derived from larval cholinergic neurons. Likewise, GABAergic/glycinergic neurons in the middle part of the cerebral ganglion were derived from larval ependymal cells (Supplementary Fig. 5). Glutamatergic neurons were not examined because only a few of them were found in the cerebral ganglion of juveniles immediately after metamorphosis. We therefore concluded that an abundance of larval ependymal cells are maintained and that some of them differentiate into adult neurons.

Furthermore, in light of a previous report suggesting that the cerebral ganglion is formed from the neurohypophysial duct[7,8], we conducted tracing experiments on this region. We introduced a promoter of *Ci-otp*, a gene encoding a homeodomain transcription factor that is expressed in the anterior part of the larval sensory vesicle including the neurohypophysial duct[21], and a Kaede fusion construct into *Ciona* embryos. Larvae that expressed Kaede exclusively in the neurohypophysial duct were selected to trace this specific region (Supplementary Fig. 6a). We found red-labelled larval cells located in the anterior tip of

the ciliated funnel (Supplementary Fig. 6b), whereas no red-coloured cells were present in the cerebral ganglion. This result suggests that the cells in the neurohypophysial duct comprise the anterior tip of the ciliated funnel, but they are not a major source of cells for the adult CNS.

Based on these results, we conclude that the anterior parts of the ascidian larval CNS, namely the sensory vesicle, posterior brain, neck and visceral ganglion, contribute to the formation of the adult CNS (Fig. 3). Thus components of the larval and adult CNS are maintained even though their body plan is subjected to extensive rearrangement during metamorphosis to form sessile adults in ascidians. In addition, the anterior–posterior axis of the larval CNS is also inherited to form the adult CNS. This indicates that the anterior–posterior axis of the adult CNS is already determined by developmental regulatory genes that are known to play essential roles in anterior–posterior axis formation of the larval CNS[6,22]. The conservative mode of adult CNS formation is completely different from the destructive mode seen in lower deuterostomes. Rather, such a conservative system is observed in vertebrate metamorphosis[23], suggesting that urochordates and vertebrates share a similar system of adult CNS formation.

In contrast to the overall maintenance of the CNS during metamorphosis, most larval neurons, except for a few motor neurons and glutamatergic neurons, disappear during the process of adult CNS formation (Fig. 3b). Most adult CNS cells are supplied from larval ependymal cells (Fig. 3a). In addition, a subset of adult neurons unambiguously originates from larval ependymal cells. Thus ascidian larval ependymal cells can act as neural stem-like cells that can give rise to several types of adult neuron during metamorphosis. In support of this idea, it has been shown that ependymal cells in vertebrates can act as neural stem cells during recovery from injury and that they also generate olfactory bulb neurons[24,25]. However, it is still controversial whether ependymal cells act as neural stem cells[26]. Indeed, the role of ependymal cells in the chordate CNS in general warrants further investigation[3,4].
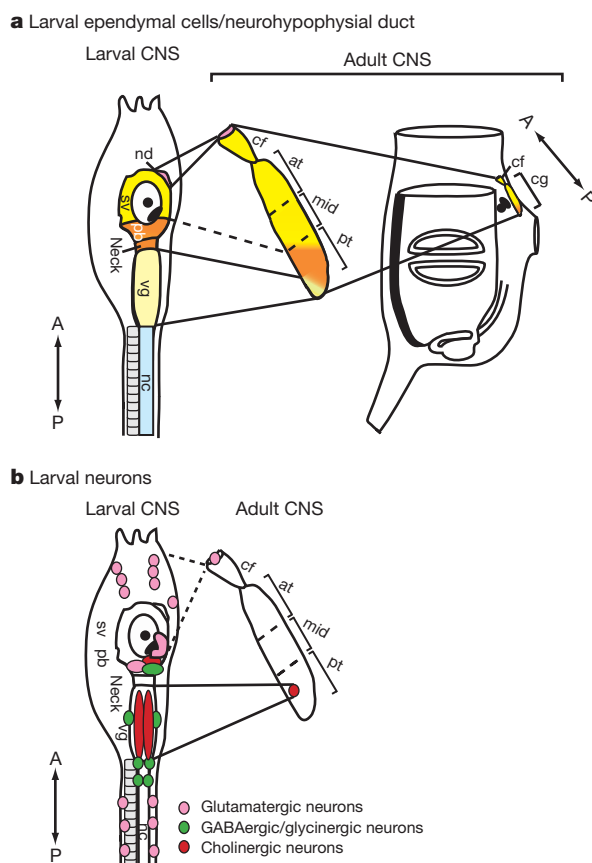
METHODS SUMMARY

Figure 3 | Adult CNS is constructed from the larval CNS. The relationship between the larval and adult CNS with respect to larval ependymal cells (**a**) and larval neurons (**b**); nd, neurohypophysial duct; sv, sensory vesicle; pb, posterior brain; vg, visceral ganglion; nc, nerve cord; cf, ciliated funnel; at, mid and pt, anterior, middle and posterior parts of the cerebral ganglion, respectively; cg, cerebral ganglion; A, anterior; P, posterior.

1. Nielsen, C. Larval and adult brains. *Evol. Dev.* **7,** 483–489 (2005).
2. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H. & Miyawaki, A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl Acad. Sci. USA* **99,** 12651–12656 (2002).
3. Barres, B. A. The mystery and magic of glia: a perspective on their roles in health and disease. *Neuron* **60,** 430–440 (2008).
4. Allen, N. J. & Barres, B. A. Glia – more than just brain glue. *Nature* **457,** 675–677 (2009).
5. Satoh, N. *Developmental Biology of Ascidians* (Cambridge University Press, 1994).
6. Wada, H., Saiga, H., Satoh, N. & Holland, P. W. Tripartite organization of the ancestral chordate brain: the antiquity of placodes: insight from ascidian *Pax-2/5/8, Hox* and *Otx* genes. *Development* **125,** 1113–1122 (1998).
7. Manni, L. *et al.* Neurogenic and non-neurogenic placodes in ascidians. *J. Exp. Zool. B* **302,** 483–504 (2004).
8. Manni, L., Agnoletto, A., Zaniolo, G. & Burighel, P. Stomodeal and neurohypophysial placodes in *Ciona intestinalis*: insights into the origin of the pituitary gland. *J. Exp. Zool. B* **304,** 324–339 (2005).
9. Dufour, H. D. *et al.* Precraniate origin of cranial motoneurons. *Proc. Natl Acad. Sci. USA* **103,** 8727–8732 (2006).
10. Kusakabe, T., Yoshida, R., Ikeda, Y. & Tsuda, M. Computational discovery of DNA motifs associated with cell type-specific gene expression in *Ciona. Dev. Biol.* **276,** 563–580 (2004).
11. Tarallo, R. & Sordino, P. Time course of programmed cell death in *Ciona intestinalis* in relation to mitotic activity and MAPK signaling. *Dev. Dyn.* **230,** 251–262 (2004).
12. Katz, M. J. Comparative anatomy of the tunicate tadpole *Ciona intestinalis. Biol. Bull.* **164,** 1–27 (1983).

13. Nicol, D. & Meinertzhagen, I. A. Cell counts and maps in the larval central nervous system of the ascidian *Ciona intestinalis* (L.). *J. Comp. Neurol.* **309,** 415–429 (1991).
14. Horie, T., Kusakabe, T. & Tsuda, M. Glutamatergic networks in the *Ciona intestinalis* larva. *J. Comp. Neurol.* **508,** 249–263 (2008).
15. Horie, T., Nakagawa, M., Sasakura, Y. & Kusakabe, T. G. Cell type and function of neurons in the ascidian nervous system. *Dev. Growth Differ.* **51,** 207–220 (2009).
16. Horie, T., Nakagawa, M., Sasakura, Y., Kusakabe, T. G. & Tsuda, M. Simple motor system of the ascidian larva: neuronal complex comprising putative cholinergic and GABAergic/glycinergic neurons. *Zool. Sci.* **27,** 181–190 (2010).
17. Takamura, K., Egawa, T., Ohnishi, S., Okada, T. & Fukuoka, T. Developmental expression of ascidian neurotransmitter synthesis genes. I. Choline acetyltransferase and acetylcholine transporter genes. *Dev. Genes Evol.* **212,** 50–53 (2002).
18. Yoshida, R. *et al.* Identification of neuron-specific promoters in *Ciona intestinalis*. *Genesis* **39,** 130–140 (2004).
19. Konno, A. *et al.* Distribution and structural diversity of cilia in tadpole larvae of the ascidian *Ciona intestinalis*. *Dev. Biol.* **337,** 42–62 (2010).
20. Tsuda, M. *et al.* Origin of the vertebrate visual cycle: II. Visual cycle proteins are localized in whole brain including photoreceptor cells of a primitive chordate. *Vision Res.* **43,** 3045–3053 (2003).
21. Moret, F. *et al.* Regulatory gene expressions in the ascidian ventral sensory vesicle: evolutionary relationships with the vertebrate hypothalamus. *Dev. Biol.* **277,** 567–579 (2005).
22. Imai, K. S., Stolfi, A., Levine, M. & Satou, Y. Gene regulatory networks underlying the compartmentalization of the *Ciona* central nervous system. *Development* **136,** 285–293 (2009).
23. Gilbert, S. F. *Developmental Biology* 8th edn (Sinauer Associates, 2006).
24. Johansson, C. F. *et al.* Identification of a neural stem cell in the adult mammalian central nervous system. *Cell* **96,** 25–34 (1999).
25. Coskun, V. *et al.* CD133$^+$ neural stem cells in the ependyma of mammalian postnatal forebrain. *Proc. Natl Acad. Sci. USA* **105,** 1026–1031 (2008).
26. Chojnacki, A. K., Mak, G. K. & Weiss, S. Identity crisis for adult periventricular neural stem cells: subventricular zone astrocytes, ependymal cells or both? *Nature Rev. Neurosci.* **10,** 153–163 (2009).
27. Sasakura, Y., Awazu, S., Chiba, S. & Satoh, N. Germ-line transgenesis of the Tc1/*mariner* superfamily transposon *Minos* in *Ciona intestinalis*. *Proc. Natl Acad. Sci. USA* **100,** 7726–7730 (2003).

Author Contributions T.H. and Y.S. designed the experiments. T.H., R.S. and Y.O. performed most of the experiments. T.H. and T.G.K. isolated *cis* regulatory elements. T.G.K. and N.S. were advisors for the experiments and evaluated the data. T.H., N.S. and Y.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.H. (horie@kurofune.shimoda.tsukuba.ac.jp).

METHODS

Biological materials. Wild-type *Ciona intestinalis* were collected from or cultivated in Onagawa Bay (Miyagi), Maizuru Bay (Kyoto) and Usa Bay (Kochi). Sperm and eggs were collected by dissecting the sperm and gonadal ducts. Transgenic lines[27] were cultured by an island system[28].

Constructs. To generate pSPCiβ2tubulin(TB)K, pSPCiVGLUTK, pSPCiVGATK pSPCiVACHTK, pSPCiPhox2K, pSPCiopsin3K, pSPCiCRALBPK and pSPCiotpK, the 5′ upstream region of *Ci-β2tubulin*, *CiPhox2*, *Ci-VGLUT*, *Ci-VGAT*, *Ci-VACHT*, *Ci-opsin3*, *Ci-CRALBP* and *Ci-otp*[9,10,14,18] were amplified by polymerase chain reaction (PCR) using thermostable DNA polymerases and gene-specific oligonucleotide primers (5′-ACGTGGATCCGATCAAGCACTGAGG GTGCT-3′ and 5′-GGGGATCCATGATGAATAGAAACTAAAGAT-3′ for *Ci-β2tubulin*; 5′-AGGGATCCCCGAAAAACAACTGCTTCTG-3′ and 5′-CCGGA TCCATCGGTGGCTGGCAGTTTC-3′ for *CiPhox2*; 5′-GATCGGATCCCCG GTATGTCCACAGCATTC-3′ and 5′-GGGGATCCAATATCCTAAATACCC TCC-3′ for *Ci-VGLUT*; 5′-TATAGGATCCTGATGGAAGATGGGACAC-3′ and 5′-TGCGGATCCACTAACAGAAGGTACCTATTC-3′ for *Ci-VGAT*; 5′-TATGGGATCCAGGCTTAAGCACACGTTC-3′ and 5′-GCCGGGATCCG ATGAACAATAAAGTAGA-3′ for *Ci-VACHT*; 5′-AGGGATCCGTTGTTTG TACCAATGTGAG-3′ and 5′-CCGGATCCATCTTGAAAATGTGTCTTCT-3′ for *Ci-opsin3*; 5′-AGGGATCCTATTTGAATTACAGTTTAAA-3′ and 5′-CCG GATCCATTGCAGATGTCGTCTGTGTA-3′ for *Ci-CRALBP*; and 5′-AGGGA TCCTTCCTGAAATGCCGCTTC-3′ and 5′-GGGGATCCATCTTTACTTAA AAAACT-3′ for *Ci-otp*). The PCR products were digested with BamHI and inserted in the BamHI site of pSP-Kaede[29]. To generate pMiCiβ2TBK, pMiCiVGLUTK, pMiCiVGATK and pMiCiVACHTK, promoter-Kaede cassettes were amplified by PCR using a thermostable DNA polymerase and vector-specific oligonucleotide primers (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTG AACTCGAGCAGCTGAAGCTTG-3′ and 5′-GGGGACCACTTTGTACAAGA AAGCTGGGTGCAGATCTGATGGCCGCTTTGAC-3′). The PCR products were subcloned into pMiDestF[29] with the Gateway system (Invitrogen). To generate pSPCiVACHTC and pSPCiVGATC, 5′ upstream regions of *Ci-VACHT* and *Ci-VGAT* were amplified with primers described above, digested with BamHI and inserted into the BamHI site of pSPCFP. To generate pSPCiVACHTCCiCRA-LBPK, the CiVACHT-CFP cassette was amplified with vector-specific primers (5′-GAACTCGAGCAGCTGAAGCTTG-3′ and 5′-GCAGATCTGATGGCCGC TTTGAC-3′), digested with HindIII and inserted into HindIII and blunted PstI sites of pSPCiCRALBPK. To generate pSPCiVGATCCiCRALBPK, the CiVGAT-CFP cassette was amplified with vector-specific primers, digested with XhoI and inserted into XhoI and blunted PstI sites of pSPCiCRALBPK.

Generation of Kaede transgenic lines. Kaede transgenic lines were created by co-electroporation of 80 μg of *in vitro* synthesized transposase messenger RNA (mRNA) and 60 μg of pMiCiβ2TBK, pMiCiVGLUTK, pMiCiVGATK and pMiCiVACHTK constructs[27,30]. The specificities of the expression of Kaede in the glutamatergic neurons, GABAergic/glycinergic neurons and cholinergic neurons in the transgenic lines were confirmed by double-fluorescence *in situ* hybridization of *Kaede* mRNA and *Ci-VGLUT*, *Ci-VGAT* and *Ci-VACHT* mRNAs, respectively. For synthesis of digoxigenin-labelled RNA probes, complementary DNA (cDNA) fragments of *Ci-VGLUT*, *Ci-VGAT* and *Ci-VACHT* were obtained from cDNA in the larval stage by reverse transcription PCR with gene-specific primers (5′-TATGGCGGCCGCCCTCCCTTCCAGTATGGG-3′ and 5′-TTAATGCTTGTAATATTTCTCATCTTCC-3′ for *Ci-VGLUT*; 5′-TATGGC GGCCGCCGAAAGACGTGGTCACC-3′ and 5′-TTAACTATAGTGCTCGAT GCTCTGTC-3′ for *Ci-VGAT*; 5′-TATGGCGGCCGCCGTTCCTGCCCCATCT TT-3′ and 5′-CTATTTTCGTTGGTATGTATGTCC-3′ for *Ci-VACHT*). The cDNA fragments were cloned into the pBluescript SK II Vector (Stratagene) and used as templates for the synthesis of the digoxigenin-labelled antisense RNA probe. Fluorescence *in situ* hybridization was performed as described previously[31]. The photoconversion of Kaede was performed by illuminating larvae with an ultraviolet light using a fluorescence microscope (Axio10, Carl Zeiss). The illuminated larvae were cultured under dark conditions until observation.

DiI labelling. DiI labelling of the A9.15 blastomere, which gives rise to ependymal cells of the visceral ganglion[32], was performed as described previously[33]. DiI (Celltracker CM-DiI, Molecular Probes) was dissolved in soybean oil at a concentration of 5 mg ml^{-1}. DiI-labelled animals were observed with a fluorescent microscope (Axio10, Carl Zeiss).

28. Joly, J. S. *et al.* Culture of *Ciona intestinalis* in closed systems. *Dev. Dyn.* **236**, 1832–1840 (2007).
29. Hozumi, A. *et al.* Efficient transposition of a single *Minos* transposon copy in the genome of the ascidian *Ciona intestinalis* with a transgenic line expressing transposase in the egg. *Dev. Dyn.* **239**, 1076–1088 (2010).
30. Matsuoka, T., Awazu, S., Shoguchi, E., Satoh, N. & Sasakura, Y. Germline transgenesis of the ascidian *Ciona intestinalis* by electroporation. *Genesis* **41**, 67–72 (2005).
31. Ikuta, T. & Saiga, H. Dynamic change in the expression of developmental genes in the ascidian central nervous system: revisit to the tripartite model and the origin of midbrain–hindbrain boundary region. *Dev. Biol.* **312**, 631–643 (2007).
32. Cole, A. G. & Meinertzhagen, I. A. The central nervous system of the ascidian larva: mitotic history of cells forming the neural tube in late embryonic *Ciona intestinalis*. *Dev. Biol.* **271**, 239–262 (2004).
33. Satou, Y., Imai, K. S. & Satoh, N. The ascidian *Mesp* gene specifies heart precursor cells. *Development* **131**, 2533–2541 (2004).

LETTER

Hydrostatic pressure and the actomyosin cortex drive mitotic cell rounding

Martin P. Stewart[1,2], Jonne Helenius[1], Yusuke Toyoda[3], Subramanian P. Ramanathan[1], Daniel J. Muller[1] & Anthony A. Hyman[3]

During mitosis, adherent animal cells undergo a drastic shape change, from essentially flat to round[1-3]. Mitotic cell rounding is thought to facilitate organization within the mitotic cell and be necessary for the geometric requirements of division[4-7]. However, the forces that drive this shape change remain poorly understood in the presence of external impediments, such as a tissue environment[2]. Here we use cantilevers to track cell rounding force and volume. We show that cells have an outward rounding force, which increases as cells enter mitosis. We find that this mitotic rounding force depends both on the actomyosin cytoskeleton and the cells' ability to regulate osmolarity. The rounding force itself is generated by an osmotic pressure. However, the actomyosin cortex is required to maintain this rounding force against external impediments. Instantaneous disruption of the actomyosin cortex leads to volume increase, and stimulation of actomyosin contraction leads to volume decrease. These results show that in cells, osmotic pressure is balanced by inwardly directed actomyosin cortex contraction. Thus, by locally modulating actomyosin-cortex-dependent surface tension and globally regulating osmotic pressure, cells can control their volume, shape and mechanical properties.

To analyse cell shape during mitosis, we simultaneously used atomic force microscopy (AFM), to measure cell height, and transmitted light microscopy, to measure cell width (Methods and Supplementary Fig. 1). Because we can determine the position of the cantilever with nanometre precision, this provides a similarly precise measure of the cell dimensions. Metaphase HeLa cells had a height-to-width ratio of 0.86 ± 0.04 (mean $\pm$ s.d.; Supplementary Fig. 1b). Mitotic cells without retraction fibres were almost spherical, as were interphase cells detached with trypsin (Supplementary Fig. 1b, c). Therefore, we conclude that a detached, isolated cell will be nearly spherical, independent of its cell cycle phase. This suggests that loss of adhesion as cells enter mitosis permits cell rounding[3].

A role for actin-based processes has previously been demonstrated in mitotic cell rounding[1,4,6,8,9]. Therefore, we tested the role of the actin cytoskeleton in maintaining a spherical shape by adding cytochalasin D to rounded cells (Supplementary Fig. 1a, e). After treatment, both detached mitotic and interphase cells remained round. However, if retraction fibres were present, rounded cells sagged to height-to-width ratios of <0.5 on cytochalasin D treatment. Therefore, the actomyosin cytoskeleton is necessary for generating a rounding force against adhesion.

To quantify the force of cell rounding, a tipless cantilever was positioned over a prophase HeLa cell, 8 μm above the substrate (Fig. 1a), and held there while the cell underwent mitosis. We refer to this method as a 'constant-height assay'. When becoming rounder in prometaphase, the mitotic cell came in contact with the cantilever and the upward force that it exerted on the cantilever was measured with subnanonewton accuracy. Simultaneously, the cell's progression through mitosis was monitored using light microscopy (Fig. 1b). Within ~10 min after nuclear envelope breakdown, cells were cylindrical, and remained so until division. As cells progressed through

prometaphase and into metaphase, the force exerted on the cantilever increased. Because cortical tension was uniform across the cell until anaphase[10] (Supplementary Fig. 2), we were able to normalize force by
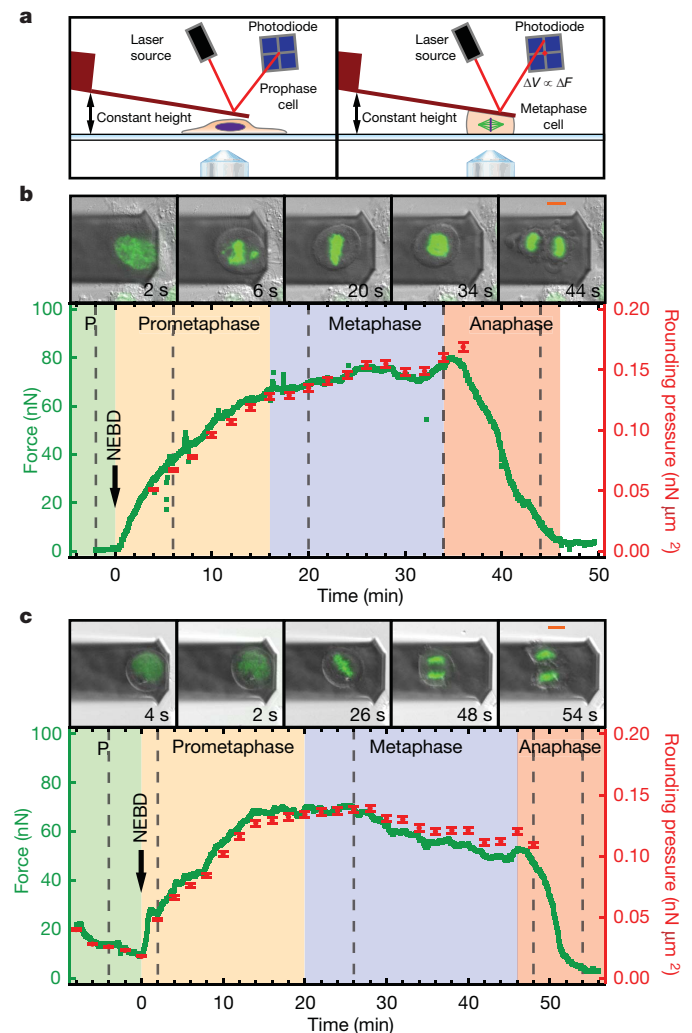

Figure 1 | Cells exert an increased rounding pressure in mitosis.
a, Constant-height assay (Methods Summary). *V*, voltage signal at the AFM photodiode; *F*, force. **b**, Overlaid differential interference contrast (DIC) and histone H2B/green fluorescent protein (GFP) images of a mitotic HeLa cell at the times indicated by the grey dashed lines. Graphed is the measured upward force (green) and calculated rounding pressure (red), which could be derived only while the cell was cylindrical (Methods). Time zero denotes nuclear envelope breakdown (NEBD). Mitotic phases are as follows: prophase (P, green), prometaphase (orange), metaphase (blue) and anaphase (red). **c**, As in **b** but for a mitotic cell pre-rounded with trypsin treatment before NEBD. Error bars, ±2% (based on measurement uncertainty from DIC images); scale bars, 10 μm.

[1]ETH Zürich, Department of Biosystems Science and Engineering, CH-4058 Basel, Switzerland. [2]Biotechnology Center, University of Technology Dresden, D-01307 Dresden, Germany. [3]Max-Planck-Institute of Molecular Cell Biology and Genetics, D-1307 Dresden, Germany.

dividing it by the cross-sectional area of the cell, providing a cell 'rounding pressure'. In metaphase, the rounding pressure reached a maximum at $0.14 \pm 0.04\,\mathrm{nN\,\mu m^{-2}}$ ($n = 83$). At anaphase, before ingression of the cleavage furrow, there was a transient rise in the rounding pressure. Concomitant with cytokinesis, force decreased, and the daughter cells flattened and finally lost contact with the cantilever. We also showed that individual G2/prophase cells that had been pre-rounded with trypsin and progressed into mitosis (Fig. 1c) increased their rounding pressure by more than 3-fold, despite cell size remaining relatively constant (Supplementary Fig. 3). Thus, as cells enter mitosis, they can exert considerable forces against external objects. These changes in mechanical properties resemble those of early studies on sea urchin eggs[11].

To determine what mechanisms generate the increased rounding force during mitosis, we examined the forces exerted by mitotic cells in the presence of various perturbants of the cytoskeleton. All actomyosin inhibitors tested significantly reduced the rounding pressure of mitotic cells (Fig. 2a). In contrast, perturbing microtubule dynamics increased the exerted pressure (Fig. 2a), perhaps because Rho/Rac guanine nucleotide exchange factor 2 was no longer inhibited[12]. Therefore, an intact actomyosin cortex, but not the microtubule cytoskeleton, is required for rounding cells to generate a rounding pressure against an external impediment.

While performing the constant-height assay on mitotic cells using intermediate concentrations of latrunculin A (40–100 nM), we noticed oscillations in the rounding force, which correlated with blebbing (Fig. 2b). The rounding pressure decreased while blebs expanded, and recovered during bleb retraction (Fig. 2b). A bleb forms when a section of membrane detaches from the actomyosin cortex[13] and retracts when the actomyosin cortex reassembles underneath the membrane and pulls the bleb back into the main cell body[14]. The concurrence of bleb formation and the decrease in rounding pressure suggests that the cell was under hydrostatic pressure: the pressure inside the cells pushing the cantilever upward was partially released when a bleb formed[15]. This interpretation is supported by recent measurements quantitatively relating cortical actomyosin tension with bleb formation[16].

The force on the cantilever could be a result of osmotic pressure. If the osmolarity is higher inside the cell than outside, water will flow into the cell and generate a hydrostatic pressure. To test this idea, we modulated the osmolarity of the medium. Introduction of hypotonic medium ($-\Delta 100\,\mathrm{mosM\,l^{-1}}$) led to an immediate increase in the volume of metaphase cells ($40 \pm 6\%$; $n = 9$), indicating that water entered the cells (Fig. 3a). This was accompanied by a concurrent increase in the measured rounding pressure ($76 \pm 20\%$; $n = 9$), presumably because the intracellular pressure increased. Within 3 min of the osmolarity changing, the cell volume and rounding pressure returned to close to their original values. This is probably because, in response to increased osmotic pressure, regulatory volume decrease causes cells to release ions[17]. Conversely, when hypertonic medium ($+\Delta 200\,\mathrm{mosM\,l^{-1}}$) was introduced (Fig. 3b) the changes in volume ($-24 \pm 4\%$; $n = 9$) and rounding pressure ($-30 \pm 14\%$; $n = 9$) were in the opposite direction. Again, the cells recovered the original rounding pressure and volume, presumably because regulatory volume increase triggers the influx of osmolytes[17].

Because ion transporters at the plasma membrane increase intracellular osmotic pressure and restore the volume of cells immediately after hypertonic challenge, we reasoned that they might also contribute to the increased rounding pressure seen in mitosis (Fig. 1). Therefore, we tested the effect of inhibiting ion transporters important in regulatory volume increase[17]. Among the inhibitors tested, an inhibitor of Na^+/H^+ antiporters, ethylisopropylamiloride, caused the greatest decrease in rounding pressure ($-53 \pm 10\%$; $n = 19$) and volume ($-8 \pm 2\%$; $n = 19$) (Fig. 3c). The exchange of a proton with a Na^+ ion increases the intracellular osmolarity because pH is strongly buffered in the cytoplasm; thus, a Na^+ ion has a greater effect on osmolarity than a
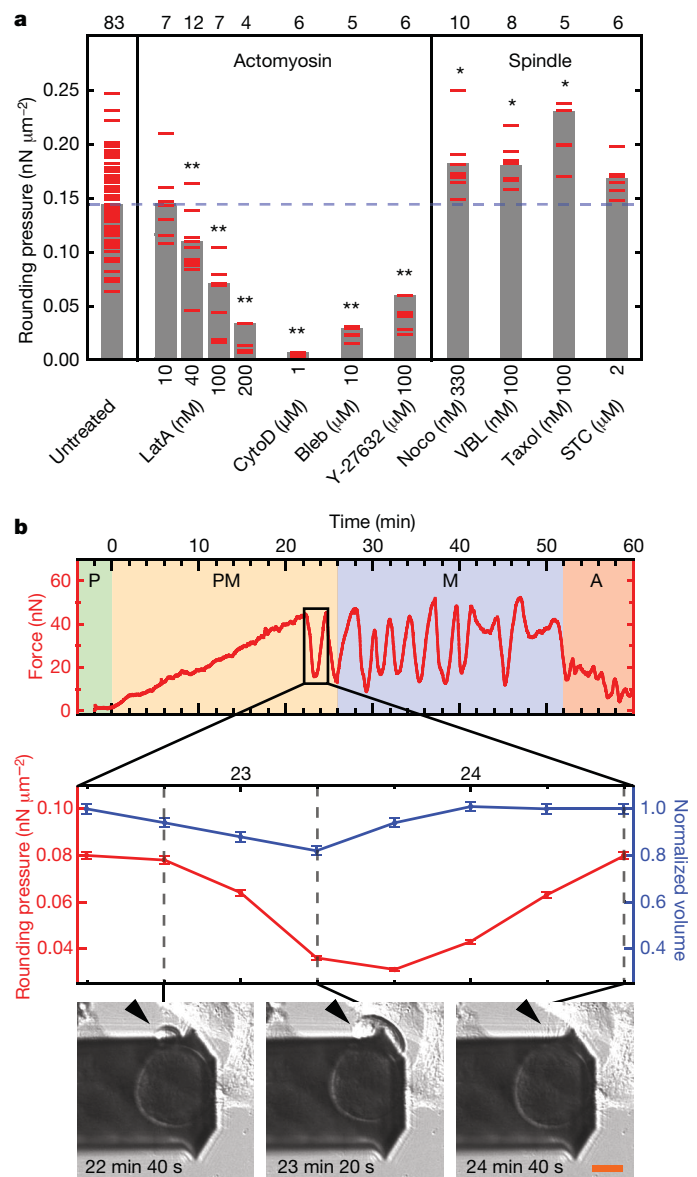

Figure 2 | Mitotic cells require a functional actin cytoskeleton to generate rounding pressure. a, Maximum rounding pressures generated by mitotic cells while incubated with inhibitors of the actomyosin system (latrunculin A (LatA, $n = 7$, 12, 7, and 4, respectively), cytochalasin D (CytoD, $n = 6$), blebbistatin (Bleb, $n = 5$), and Y-27632 ($n = 6$)) or microtubule spindle perturbants (nocodazole (Noco, $n = 10$), vinblastine (VBL, $n = 8$), Taxol (TXL, $n = 5$) and S-trityl-cysteine (STC, $n = 6$)). Red marks are maximum rounding pressures generated by single cells. Grey bars denote averages. n values are shown above each category. *$P < 0.05$, **$P < 0.001$. **b**, Top: rounding force exerted by a mitotic cell incubated with 40 nM LatA. Bottom: rounding pressure (red) and cell volume (not including the bleb, blue) during an oscillation, with corresponding DIC images showing bleb expansion and retraction. The drop in rounding pressure ($< -50\%$) cannot be accounted for by the volume decrease in the main cell body ($\sim -15\%$) indicating a decrease in intracellular pressure. A, anaphase; M, metaphase; PM, prometaphase. Error bars, $\pm 2\%$ (based on measurement uncertainty from DIC images); scale bar, 10 μm.

proton[17,18]. These results suggest the Na^+/H^+ antiporter increases mitotic rounding pressure, and agree well with data showing that the Na^+/H^+ antiporter SLC9A1 is activated at the G2–M transition[19].

We next perturbed the ion gradients across the plasma membrane using *Staphylococcus aureus* α-toxin, which confers permeability to monovalent cations[20]. When added to metaphase cells, it caused a decrease in both volume ($-42 \pm 4\%$; $n = 11$) and rounding pressure
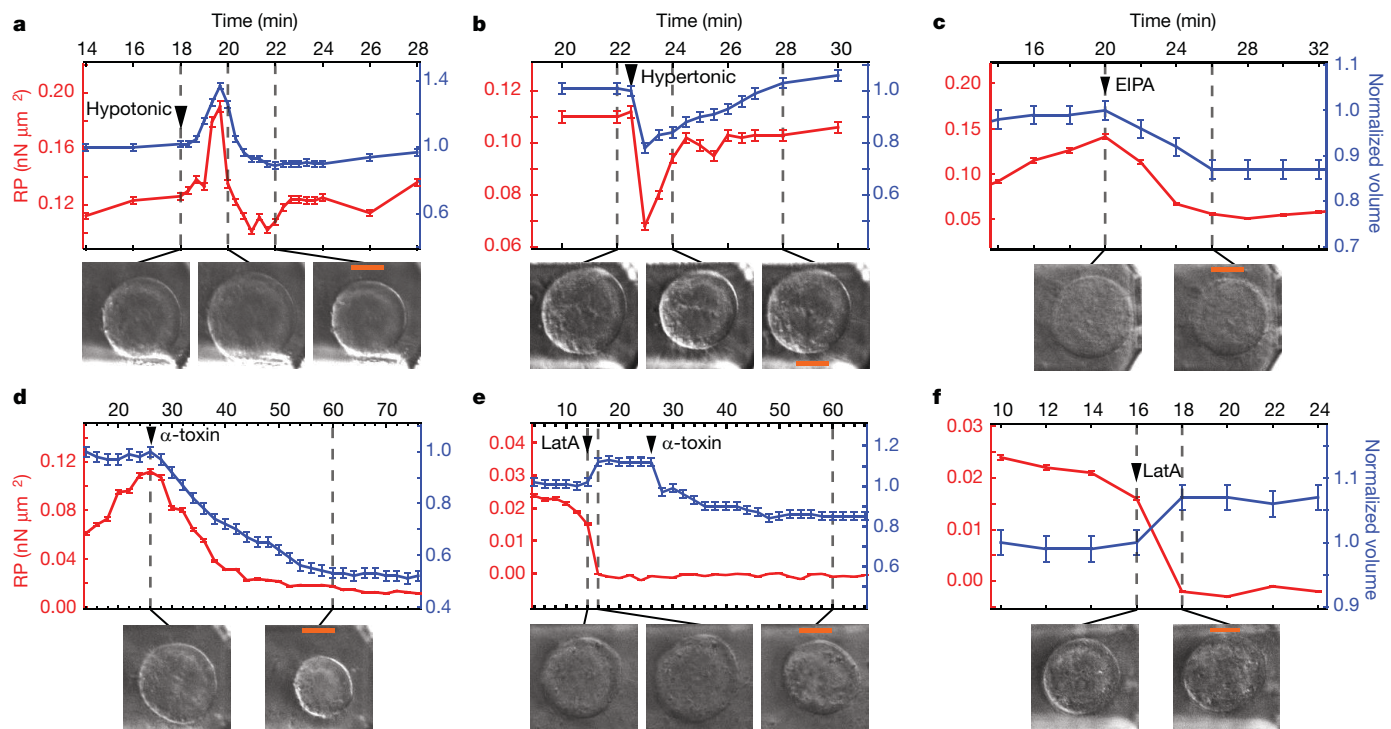
Figure 3 | The actomyosin cortex contracts against an intracellular osmotic pressure. Representative rounding pressure (RP) and cell volume time courses for mitotic cells subjected to the following perturbations: hypotonic ($-\Delta100\ \text{mosM}\,\text{l}^{-1}$) medium ($n = 9$) (**a**); hypertonic ($+\Delta200\ \text{mosM}\,\text{l}^{-1}$; $+3\%$ xylose) medium ($n = 9$) (**b**); $50\ \mu\text{M}$ ethylisopropylamiloride (EIPA, $n = 19$) (**c**); $60\ \mu\text{g}\ \text{ml}^{-1}$ α-toxin ($n = 11$) (**d**); $1\ \mu\text{M}$ latrunculin A (14 min) then α-toxin (26 min, $n = 9$) (**e**); and $1\ \mu\text{M}$ latrunculin A ($n = 36$) (**f**). In **e** and **f**, pre-treatment with $5\ \mu\text{M}$ blebbistatin prevented latrunculin-A-induced blebbing. Time zero represents NEBD. Error bars, $\pm2\%$ (based on measurement uncertainty from DIC images); scale bars, $10\ \mu\text{m}$.

($-88 \pm 8\%$; $n = 11$) (Fig. 3d). Another pore-forming toxin, *Escherichia coli* haemolysin A (HlyA), which also renders the plasma membrane permeable to cations[21], also reduced the volume and rounding pressure of mitotic cells (Supplementary Fig. 4). Disruption of the actomyosin cortex with a combination of blebbistatin and latrunculin A before *S. aureus* α-toxin treatment reduced the toxin's ability to decrease cell volume (Fig. 3e), suggesting a contribution of the actomyosin cortex. We do not know whether this was due to an inward pressure caused by the actomyosin cortex or to more indirect mechanisms of volume regulation.

To further probe the link between osmotic pressure and actomyosin contraction, we performed experiments where we spontaneously abolished or stimulated the actomyosin cortex. If there is an intracellular pressure opposed by the actomyosin cortex, disruption of the cortex should result in dissipation of intracellular pressure and a small increase in cell volume. Indeed, when we treated mitotic cells first with blebbistatin, which inhibits myosin II contraction, and then with latrunculin A to depolymerize actin filaments, the volume of mitotic cells increased by $7 \pm 4\%$ ($n = 36$) and the mitotic rounding pressure was abolished (Fig. 3f). To study the converse case, we then looked at the effects of instantaneous activation of the actomyosin cortex. To do this we took advantage of blebbistatin's propensity to be inactivated by blue light[22]. When we photoinactivated blebbistatin, mitotic cells responded with an increase in rounding pressure and a decrease in volume (Supplementary Fig. 5). We conclude that stimulating contraction of the actomyosin cytoskeleton increases rounding pressure and decreases volume, whereas disrupting actomyosin activity reduces rounding pressure and increases volume.

Our experiments show that perturbation of osmotic gradients, associated transporters and the actomyosin cortex caused changes in both volume and rounding pressure (Fig. 4). When osmotic pressure was reduced, rounding pressure and volume decreased (Fig. 4a,

b, lower left quadrants). Conversely, if osmotic pressure was increased, rounding pressure and volume increased (Fig. 4a, b, upper right quadrants). When the actomyosin cytoskeleton was abolished, cell volume increased while rounding pressure decreased (Fig. 4a, b, lower right quadrants). However, rounding pressure increased as volume decreased after actomyosin contraction was stimulated (Fig. 4a, b, upper left quadrants). In conclusion, these results reveal that the actomyosin cortex contracts against an opposing intracellular osmotic pressure.

We propose the following model for the active processes that drive cell rounding during mitosis. As cells enter mitosis, de-adhesion from the substrate allows cells to become rounder[3] (Supplementary Fig. 1). At the same time, cells increase their intracellular pressure (Fig. 1), presumably to drive rounding in a tightly packed tissue environment. Our model implies that intracellular pressure is at least equal to the measured rounding pressure, $\sim150\ \text{Pa}$ ($0.15\ \text{nN}\ \mu\text{m}^{-2}$; Fig. 1). A pressure difference across the cell membrane of 100–500 Pa is thought to be sufficient to cause cell blebbing[15] and is within the range typically measured in micropipette aspiration techniques[23] (1–1,000 Pa). A corollary to this model is that a non-homogeneous cortex results in dissimilar cell surface curvatures such as those observed in blebbing cells. In conjunction with an intracellular hydrostatic pressure, local modulation of cortical tension would allow cells to alter their shape[24,25], control their motion[25–29] and govern the mechanics of mitosis[15].

Pressure gradients are known to drive shape changes in organisms with cell walls[30]. Our experiments support the idea that the actomyosin cortex behaves like an internal cell wall that directs osmotic expansion to control animal cell shape[25–29]. Given the intricate shapes microorganisms and plants are able to achieve using turgor pressure, it is perhaps not surprising that animal cells have also evolved a mechanism that makes use of osmotic pressure.
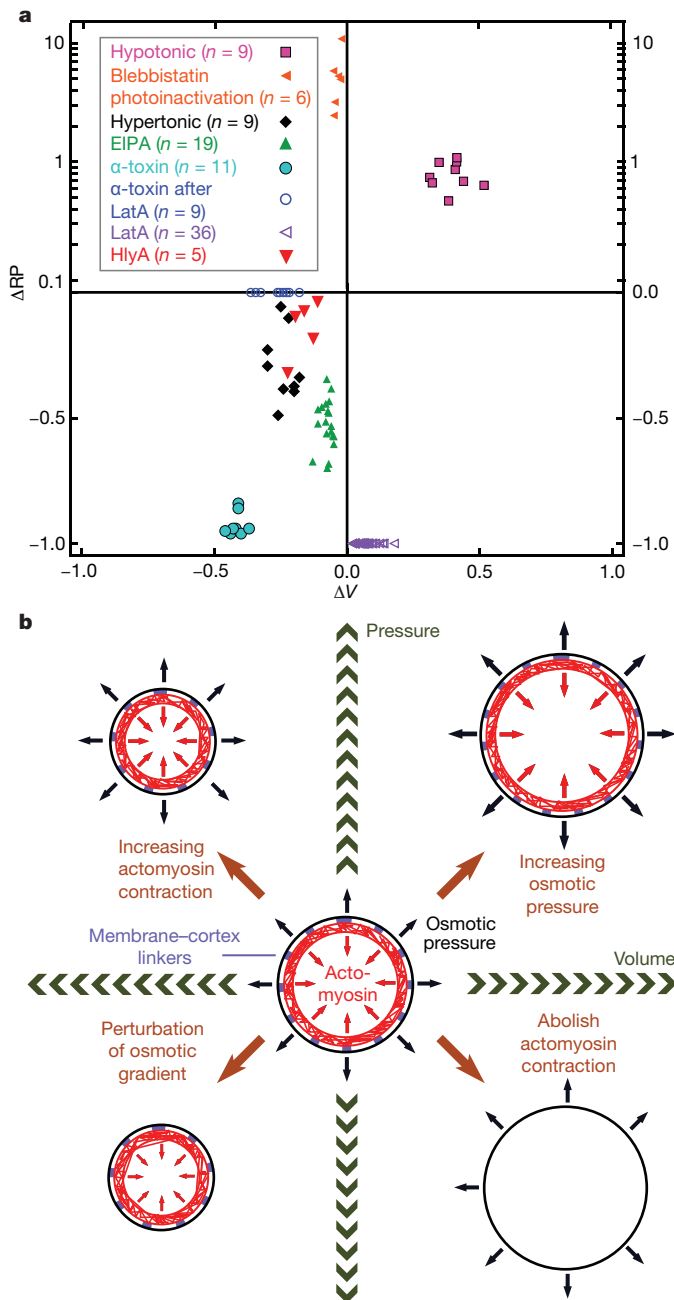
Figure 4 | Animal cells control shape in mitosis by modulating intracellular pressure in conjunction with actomyosin activity. **a**, Relative maximum changes in rounding pressure (ΔRP) and normalized volume (ΔV) of mitotic cells upon treatment with the indicated perturbations. Concentrations used are those indicated in Fig. 3 legend, 2 μg ml^{-1} for haemolysin A (HlyA) and 10 μM for blebbistatin photoinactivation, which involved a 1-s exposure to blue light. ΔRP is shown on a logarithmic scale above ΔRP $= 0.1$. *n* values are displayed in the key. **b**, Uniform actomyosin contractile tension (red) is balanced by an outward-directed, intracellular osmotic pressure (black). Membrane–cortex linkers (purple) couple these two elements. The higher the tension and pressure, the greater the cortex rigidity. Imbalances between tension and osmotic pressure cause changes in cell volume and rounding force.

METHODS SUMMARY

Cell culture. HeLa-Kyoto cells expressing a fluorescent histone construct (H2B–GFP) were grown to ~50% confluency on 24-mm-diameter glass coverslips for cantilever experiments. We used DMEM containing 4 mM sodium bicarbonate (PN:31600-083, Invitrogen) buffered with 20 mM HEPES for experiments.
Instrumentation. The experimental set-up consisted of an AFM (Nanowizard I, JPK Instruments) mounted on a Zeiss Axiovert 200M optical microscope. Tipless

cantilevers were ~250 μm long, 35 μm wide, 2 μm thick and made of pure silicon (NSC12-D/tipless/noAl, $k \approx 0.3$ N m^{-1}, Mikromasch). A BioCell (JPK Instruments) maintained cells at 37 °C.
AFM constant-height assay. We positioned the end of the cantilever over a candidate cell, 8 μm above the substrate. A laser beam was used to monitor the position of the cantilever, which was calibrated and used to record forces generated by cells. Rounding pressure was derived by dividing the measured force by the horizontal cross-sectional area of the cell, which was measured using DIC images. Cell volume was determined by multiplying the cross-sectional area of the near-cylindrical cell by its height under the cantilever. The ± 2% error in rounding pressure and volume shown in the figures is based on measurement uncertainty from DIC images.

In Fig. 2a, cells were pre-incubated with perturbants. In Fig 3a, b, tonic shock was induced by exchanging the full volume of the AFM BioCell several times with WPI Aladdin push–pull pumps. In Fig. 3c–f, perturbants were added to the AFM BioCell with a microsyringe.

We determined the following mitotic phases from H2B–GFP images: prophase, condensed chromosomes but intact nucleus; prometaphase, nuclear envelope breakdown; metaphase, chromosomes aligned to form a metaphase plate; anaphase, two sets of chromosomes separated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cramer, L. P. & Mitchison, T. J. Investigation of the mechanism of retraction of the cell margin and rearward flow of nodules during mitotic cell rounding. *Mol. Biol. Cell* **8**, 109–119 (1997).
2. Gibson, M. C., Patel, A. B., Nagpal, R. & Perrimon, N. The emergence of geometric order in proliferating metazoan epithelia. *Nature* **442**, 1038–1041 (2006).
3. Harris, A. Location of cellular adhesions to solid substrata. *Dev. Biol.* **35**, 97–114 (1973).
4. Carreno, S. *et al.* Moesin and its activating kinase Slik are required for cortical stability and microtubule organization in mitotic cells. *J. Cell Biol.* **180**, 739–746 (2008).
5. Kunda, P. & Baum, B. The actin cytoskeleton in spindle assembly and positioning. *Trends Cell Biol.* **19**, 174–179 (2009).
6. Kunda, P., Pelling, A. E., Liu, T. & Baum, B. Moesin controls cortical rigidity, cell rounding, and spindle morphogenesis during mitosis. *Curr. Biol.* **18**, 91–101 (2008).
7. Théry, M. & Bornens, M. Cell shape and cell division. *Curr. Opin. Cell Biol.* **18**, 648–657 (2006).
8. Fujibuchi, T. *et al.* AIP1/WDR1 supports mitotic cell rounding. *Biochem. Biophys. Res. Commun.* **327**, 268–275 (2005).
9. Maddox, A. S. & Burridge, K. RhoA is required for cortical retraction and rigidity during mitotic cell rounding. *J. Cell Biol.* **160**, 255–265 (2003).
10. Matzke, R., Jacobson, K. & Radmacher, M. Direct, high-resolution measurement of furrow stiffening during division of adherent cells. *Nature Cell Biol.* **3**, 607–610 (2001).
11. Hiramoto, Y. Mechanical properties of sea urchin eggs. I. Surface force and elastic modulus of the cell membrane. *Exp. Cell Res.* **32**, 59–75 (1963).
12. Krendel, M., Zenke, F. T. & Bokoch, G. M. Nucleotide exchange factor GEF-H1 mediates cross-talk between microtubules and the actin cytoskeleton. *Nature Cell Biol.* **4**, 294–301 (2002).
13. Cunningham, C. C. Actin polymerization and intracellular solvent flow in cell surface blebbing. *J. Cell Biol.* **129**, 1589–1599 (1995).
14. Charras, G. T., Hu, C. K., Coughlin, M. & Mitchison, T. J. Reassembly of contractile actin cortex in cell blebs. *J. Cell Biol.* **175**, 477–490 (2006).
15. Charras, G. T., Coughlin, M., Mitchison, T. J. & Mahadevan, L. Life and times of a cellular bleb. *Biophys. J.* **94**, 1836–1853 (2008).
16. Tinevez, J. Y. *et al.* Role of cortical tension in bleb growth. *Proc. Natl Acad. Sci. USA* **106**, 18581–18586 (2009).
17. Lang, F. *Mechanisms and Significance of Cell Volume Regulation* (Karger, 2006).
18. Wehner, F., Olsen, H., Tinel, H., Kinne-Saffran, E. & Kinne, R. K. Cell volume regulation: osmolytes, osmolyte transport, and signal transduction. *Rev. Physiol. Biochem. Pharmacol.* **148**, 1–80 (2003).
19. Putney, L. K. & Barber, D. L. Na-H exchange-dependent increase in intracellular pH times G2/M entry and transition. *J. Biol. Chem.* **278**, 44645–44649 (2003).
20. Valeva, A. *et al.* Staphylococcal alpha-toxin: repair of a calcium-impermeable pore in the target cell membrane. *Mol. Microbiol.* **36**, 467–476 (2000).
21. Koschinski, A. *et al.* Why *Escherichia coli* alpha-hemolysin induces calcium oscillations in mammalian cells—the pore is on its own. *FASEB J.* **20**, 973–975 (2006).
22. Kolega, J. Phototoxicity and photoinactivation of blebbistatin in UV and visible light. *Biochem. Biophys. Res. Commun.* **320**, 1020–1025 (2004).
23. Hochmuth, R. M. Micropipette aspiration of living cells. *J. Biomech.* **33**, 15–22 (2000).
24. Salbreux, G., Joanny, J. F., Prost, J. & Pullarkat, P. Shape oscillations of non-adhering fibroblast cells. *Phys. Biol.* **4**, 268–284 (2007).

25. Blaser, H. *et al.* Migration of zebrafish primordial germ cells: a role for myosin contraction and cytoplasmic flow. *Dev. Cell* **11,** 613–627 (2006).
26. Bereiter-Hahn, J. Mechanics of crawling cells. *Med. Eng. Phys.* **27,** 743–753 (2005).
27. Charras, G. & Paluch, E. Blebs lead the way: how to migrate without lamellipodia. *Nature Rev. Mol. Cell Biol.* **9,** 730–736 (2008).
28. Mitchison, T. J., Charras, G. T. & Mahadevan, L. Implications of a poroelastic cytoplasm for the dynamics of animal cell shape. *Semin. Cell. Dev. Biol.* **19,** 215–223 (2008).
29. Keren, K., Yam, P. T., Kinkhabwala, A., Mogilner, A. & Theriot, J. A. Intracellular fluid flow in rapidly moving cells. *Nature Cell Biol.* **11,** 1219–1224 (2009).
30. Harold, F. M. To shape a cell: an inquiry into the causes of morphogenesis of microorganisms. *Microbiol. Rev.* **54,** 381–431 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

METHODS

Cell culture. HeLa-Kyoto cells expressing a histone H2B–eGFP construct (H2B–GFP) were used[31]. In some experiments, cells additionally expressing mCherry–CAAX, fluorescently marking the plasma membrane, were used. To construct the mCherry–CAAX plasmid (TH0477), a DNA sequence (5′-TGCATGAGCTGC AAGTGTGTGCTGTCC-3′) from the carboxy terminus of the rat c-H-Ras 1 gene was inserted to the C terminus of the mCherry gene[32]. The mCherry–CAAX plasmid was transfected into H2B–GFP cells and selected by puromycin resistance. Cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS), 2 mM GlutaMAX, 100 units ml^{-1} penicillin, 100 µg ml^{-1} streptomycin and 0.5 mg ml^{-1} Geneticin (all Invitrogen) at 37 °C in a 5% CO_2 environment. DMEM used for AFM experiments (cat. no. 31600-083, Invitrogen) contained only 4 mM sodium bicarbonate and was instead buffered with 20 mM HEPES/NaOH to pH 7.2. When using dynasore and trypsin, NuSerum (Fischer Scientific) was substituted for FBS at the corresponding concentration. For AFM experiments, cells were plated on 24-mm-diameter glass coverslips (Marienfeld) and grown for 2 days until they reached ~50% confluency.

Perturbants. Chemical inhibitors were acquired from Sigma-Aldrich except jasplakinolide (Merck) and trypsin (Invitrogen) and used at the indicated concentrations. In cases where perturbants were added during a measurement, a microsyringe (Hamilton) was used to inject agents into the AFM BioCell or the full volume was exchanged several times with a WPI Aladdin push–pull pump set-up (Supplementary Fig. 7a). Supplementary Table 1 lists all perturbants used and how they were added to cells, and provides a short description of what the perturbant does.

Pore-forming-toxin experiments. *S. aureus* α-toxin stock was made at 2 mg ml^{-1} in PBS. Owing to having limited amounts of toxin stock, we added 100 µl of a ×5 concentrate to 400 µl of medium to arrive at the final concentration. Concentrate was added with a tube and syringe in the manner shown in Supplementary Fig. 7a. At a given concentration of α-toxin, more crowded coverslips seemed to dilute the potency of the toxin. Thus, for toxin experiments we attempted to keep the cell confluency constant at 40–50%. Concentrations at and below 20 µg ml^{-1} had no noticeable effect on cells. We found that a dosage of 40–80 µg ml^{-1} caused effects such as those seen in Fig. 3d.

Haemolysin A stock was made up at 910 µg ml^{-1} in 8 M guanine/HCl solution. Again, owing to having limited amounts of toxin stock we used the ×5 concentrate 'addition' method described for α-toxin. We found that a final concentration of 1 µg ml^{-1} had little effect on cells. However, using 2–4 µg ml^{-1} we obtained typical results seen in Supplementary Fig. 4.

Osmolarity change experiments. For osmolarity change experiments, the volume of the AFM BioCell (~400 µl) was exchanged four times at a flow rate of 2,500 µl min^{-1} using a WPI Aladdin push–pull pump. For hypotonic exchanges, DMEM (290 mosM kg^{-1}) was diluted with water to 190 mosM kg^{-1}. In the case of hypertonic exchanges, DMEM was supplemented with xylose to increase osmolarity to 490 mosM kg^{-1}. To minimize disturbance of the AFM laser, the refractive index of all solutions was matched with Ficoll (Sigma-Aldrich), an inert sucrose polymer with negligible effect on osmolarity in solution.

Mid-experiment introduction of perturbants: balancing refractive index and osmolarity difference. In cases where the perturbant mixture differed from the original medium in refractive index, we added Ficoll to the solution having lower refractive index until the solutions matched. 1% w/v Ficoll increased refractive index by 0.0013 with negligible change to osmolarity.

In cases where the perturbant mixture and original medium had an unintended mismatch in osmolarity, we used xylose to raise the osmolarity of the lower-osmolarity solution. 1% w/v xylose increases osmolarity by 70 mosM l^{-1} but also increased refractive index by 0.0013. In such cases, Ficoll was additionally used to match refractive index. In all cases, Ficoll and xylose never exceeded 3% w/v and did not negatively effect progression through mitosis or the rounding forces measured.

Instrumentation. The experimental set-up consisted of an AFM (Nanowizard I, JPK Instruments) mounted on an Axiovert 200M optical microscope (Carl Zeiss). A BioCell (JPK Instruments) allowed cells to be cultured at 37 °C during experiments. Tipless cantilevers were ~250 µm long, 35 µm wide, 2 µm thick and made of pure silicon (NSC12-D/tipless/noAl from Mikromasch) and had a nominal force constant of 0.3 N m^{-1}. Cantilevers were calibrated using the thermal noise method[33].

AFM constant-height assay. The procedure for an 8-µm constant-height assay of a mitotic HeLa cell was as follows. The height of the substrate adjacent to a cell was determined. Then the cantilever end was positioned over the cell, 8 µm above the substrate. Prophase cells typically had heights of <7 µm, and were therefore initially not in contact with the cantilever. The force exerted by the rounding mitotic cell was sensed by the cantilever and recorded over time. Rounding pressure was derived by dividing the force measured by the horizontal cross-sectional

area of the cell, which was measured from DIC images of the cell (Supplementary Fig. 7c). Cell volume was determined by multiplying the cross-sectional area of the near-cylindrical cell by its height under the cantilever (Supplementary Fig. 7c). The ±2% error on rounding pressure and volume is based on measurement uncertainty from DIC images. After a constant-height assay, the height of the substrate was re-measured to assess mechanical drift. Experiments with more than 10% drift in relative height were discarded. Cantilever deflection drift was usually <2 nN per hour and could be neglected when considering the magnitude of forces measured.

Cell height and shape measurements. Cell heights were determined using AFM. Because the cell is soft, particularly when treated with actin perturbants, the cell height was determined by extracting a contact point from force–distance curves. As the cantilever approached the cell, a small force, of ~100–200 pN, is registered at the cantilever upon contact. The vertical distance between this contact point and the substrate was defined as the cell height. To obtain a cell height-to-width ratio, the cell's width was measured using DIC microscopy images.

Optical microscopy. For optical microscopy, a Plan Apochromat ×20/0.8 objective lens (Zeiss) was used. Images were acquired with a CoolSNAP cf camera (Roper Scientific) driven by METAMORPH software (Visitron). An alternative set-up featuring an MRM camera (Zeiss) and AXIOVISION software (Zeiss) was used to acquire and analyse images in some experiments, including the ApoTome optical sectioning shown in Supplementary Fig. 7b. For H2B–GFP and mCherry–CAAX imaging, FITC and Texas red filter sets were respectively used.

Mitotic phase assignment. H2B–GFP images were used to determine the phase of mitosis according to the following criteria: prophase cells contain condensed chromosomes surrounded by a nuclear envelope; prometaphase starts with nuclear envelope breakdown; metaphase is when chromosomes align at the metaphase plate; and anaphase is when the two sets of chromosomes separate.

Cortical stiffness measurements. The measurement was made by indenting the cell cortex using a 5-µm-diameter bead glued to a NPO tipless cantilever (Veeco; nominal stiffness, 50 mN m^{-1}) with a force of 2 nN. The Young's modulus at each indentation was calculated by least-squares fitting a parabolic curve to the force–indentation curve at contact[34]. The method was adapted from ref. 35. The bead diameter was determined using its DIC image and the cantilever spring constant was determined as described above. We assumed the cell to have a Poisson's ratio of 0.5 (ref. 35).

Constant-height assay design. We started these experiments by using soft cantilevers (Nanoworld TL1, $k \approx 50$ mN m^{-1}). However they could neither measure nor apply forces over 20 nN; therefore, we changed to stiff cantilevers. We prefer them because they allowed us to measure a wide range of forces, from tens of piconewtons to hundreds of nanonewtons. Using such cantilevers in an 8-µm constant-height assay (Supplementary Fig. 7) allowed us to sample force, rounding pressure and volume continuously. This was key to understanding perturbations where volume and pressure respond quickly (Figs 2 and 3). Additionally, we compared the effect of long-term constraint under the cantilever with measurements on a shorter timescale using a simple rheological compression assay (Supplementary Fig. 8c). We plotted our two data sets, constant-height assays ($n = 123$) and compression assays ($n = 100$), as force versus deformation (Supplementary Fig. 8a) and contact stress versus strain (Supplementary Fig. 8b). Because we get similar results whether we 'constrain the cell' or make a simple mechanical measurement on an already round one, we conclude that our constant-height assay is applicable to studying mitotic cell rounding over a broad range of deformations and timescales.

Cell viability under the cantilever. To investigate the effect of constraining cell shape in our assay on progression through mitosis, we measured the times taken for individual cells to progress from prophase to anaphase (Supplementary Fig. 9a). Cells either subjected to a constant force of 1–50 nN or constrained to a constant height of 8 µm progressed through mitosis unaffected. However, addition of cytochalasin D to cells in an 8-µm constant-height assay slightly extended the time between nuclear envelope breakdown and anaphase. Additionally, the height-to-width ratio attained by mitotic cells becoming rounder against constant forces is plotted in Supplementary Fig. 9b.

31. Neumann, B. *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nature Methods* **3,** 385–390 (2006).
32. Ruta, M. *et al.* Nucleotide sequence of the two rat cellular rasH genes. *Mol. Cell. Biol.* **6,** 1706–1710 (1986).
33. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868–1873 (1993).
34. Hertz, H. Uber den Kontakt elastischer Korper. *J. Reine Angew. Math.* **92,** 156–172 (1881).
35. Radmacher, M., Fritz, M. & Hansma, P. K. Imaging soft samples with the atomic force microscope: gelatin in water and propanol. *Biophys. J.* **69,** 264–270 (1995).

LETTER

Leadership, social capital and incentives promote successful fisheries

Nicolás L. Gutiérrez[1], Ray Hilborn[1] & Omar Defeo[2]

One billion people depend on seafood as their primary source of protein and 25% of the world's total animal protein comes from fisheries[1]. Yet a third of fish stocks worldwide are overexploited or depleted[1,2]. Using individual case studies, many have argued that community-based co-management[3] should prevent the tragedy of the commons[4] because cooperative management by fishers, managers and scientists often results in sustainable fisheries[3,5,6]. However, general and multidisciplinary evaluations of co-management regimes and the conditions for social, economic and ecological success within such regimes are lacking. Here we examine 130 co-managed fisheries in a wide range of countries with different degrees of development, ecosystems, fishing sectors and type of resources. We identified strong leadership as the most important attribute contributing to success, followed by individual or community quotas, social cohesion and protected areas. Less important conditions included enforcement mechanisms, long-term management policies and life history of the resources. Fisheries were most successful when at least eight co-management attributes were present, showing a strong positive relationship between the number of these attributes and success, owing to redundancy in management regulations. Our results demonstrate the critical importance of prominent community leaders and robust social capital[7], combined with clear incentives through catch shares and conservation benefits derived from protected areas, for successfully managing aquatic resources and securing the livelihoods of communities depending on them. Our study offers hope that co-management, the only realistic solution for the majority of the world's fisheries, can solve many of the problems facing global fisheries.

Fish are a critical natural resource, yet global catches have peaked while human populations and demand for seafood continue to rise[1]. This increasing pressure has coincided with most fisheries worldwide being fully exploited or requiring rebuilding[2]. In the past several decades, researchers have examined the circumstances under which common pool resources, and fisheries in particular, can be successfully managed[3,5]. The dominant theme in fisheries management has been that privatization is necessary to avoid Hardin's tragedy of the commons[4], whereas Ostrom and others[6–9] have argued that community-based co-management can often achieve sustainability.

Community-based co-management (hereafter co-management) occurs when fishers and managers work together to improve the regulatory process. Advantages of co-management include: enhanced sense of ownership encouraging responsible fishing; greater sensitivity to local socioeconomic and ecological restraints; improved management through use of local knowledge; collective ownership by users in decision making; increased compliance with regulations through peer pressure; and better monitoring, control and surveillance by fishers[9,10].

Despite the increasingly widespread adoption of co-management for solving governance issues[11,12], few attempts have been made to synthesize individual case studies into a general fisheries co-management model. There are qualitative case studies, comparative analyses and a few localized quantitative reviews on the subject[12,13], but no comprehensive

evaluations to support the hypothesis that co-management improves fisheries' governance systems and performance indicators[14]. Here, we tested whether co-management improves fisheries' social, economic and ecological success, identified relevant attributes generated by isolated study cases in diverse disciplines (such as ecology and social sciences) and evaluated the relative merits of different co-management attributes across fisheries.

We assembled worldwide data from the peer-reviewed literature, government and non-governmental organization (NGO) reports and from interviews of experts on co-managed fisheries. We identified 130 co-managed fisheries in 44 countries (Fig. 1 and Supplementary Table 1) covering artisanal and industrial sectors, and a variety of ecosystem types, degrees of human development (Human Development Index (HDI)[15]), and social, economic and political settings (Supplementary Table 2). We extracted 19 variables relating co-management attributes under five categories suggested by Ostrom[16] for analysing social–ecological systems (Table 1 and Supplementary Table 2). These were used to predict eight binary measures of success grouped into ecological (for example, increase in stock abundance), social (for example, increase in social welfare) and economic (for example, increase in unit price) indicators and summed them to obtain a single holistic success score that captures natural and human dimensions of fisheries[17].

Statistically demonstrating a causal connection between co-management attributes and successful fisheries is challenging, because we are mostly dealing with non-experimental and observational studies in which random treatments and control groups are not present.
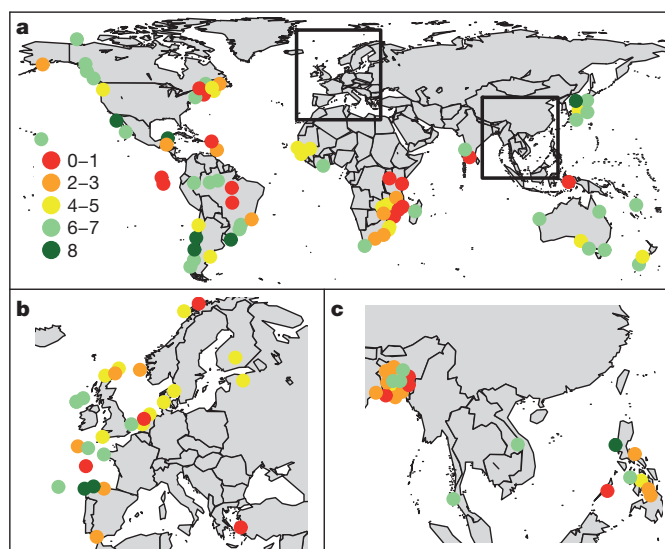

Figure 1 | Location and success score for all study cases of fisheries co-management. a–c, Success was grouped in five categories according to number of social, ecological and economic outcomes achieved. **a,** Global map. Insets are Europe (**b**) and Southeast Asia (**c**). $n = 130$.

[1]School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195-5020, USA. [2]UNDECIMAR, Facultad de Ciencias, Iguá 4225, PO Box 10773, Montevideo 11400, Uruguay.

Table 1 | Fisheries co-management attributes and outcomes.

Group	Variable name	Frequency (%)
Co-management	Type (consultative, cooperative, delegated)	-
	Phase (pre-, implementation, post-)	-
	Time frame	-
Resource system	HDI (low, medium, high, very high)	-
	Governance Index	-
	Corruption Perceptions Index	-
	Resource type (single*, multi-species)	-
	Ecosystem (inland, coastal, offshore)	-
	Fishing sector (artisanal, industrial, sequential)	-
	Defined geographic boundaries	52
Resource unit	Sedentary/low mobility resources	38
Governance system	Central government support (local)	93
	Scientific advice	92
	Minimum size restrictions	76
	Long-term management policy	71
	Global catch quotas	52
	Monitoring, control and surveillance	47
	Protected areas	39
	Spatially explicit management	37
	Individual or community quotas	33
	Co-management in law (national)	32
	Seeding or restocking programs	19
	TURF	18
Users system	Social cohesion	78
	Self-enforcement mechanisms	71
	Leadership	62
	Tradition in self-organization	55
	Influence in local market	28
Outcomes	Community empowerment	85
	Fishery status (under or fully, over-exploited)	67
	Sustainable catches	62
	Increase in social welfare	61
	Increase in catch per unit of effort	54
	Add-on conservation benefits	45
	Increase in abundance	38
	Increase in unit prices	30

All attributes were grouped according to the classification of Ostrom[16]. Values in the frequency column denote percentage of co-management attributes reported as present within the co-management systems. For complete variable descriptions see Supplementary Table 2.
* Benthic, demersal, pelagic, mammal.

However, the large number of fisheries involved in our study, covering a wide spectrum of social, ecological and political settings, and the detailed information contained in the reviewed documents, provided the basis to assess causality through several criteria: (1) strength of association between co-management attributes and success measured by robust statistical methods; (2) consistency of association in various conditions across ecosystems, fishing sectors and degrees of human development; (3) plausibility of causal explanations; (4) coherence with co-management theories and knowledge of each fishery; and (5) temporality, where presence of attributes preceded success[18]. Furthermore, although comparison to top-down management would be of interest, the objective of this study was to identify and quantify the co-management attributes determining successful fisheries, and not explicitly to compare its performance with top-down centralized management.

We tested whether success scores differed among socio-economic conditions (HDI, fishing sector) and ecological settings (ecosystems, life history of exploited resources) and we identified specific attributes associated with their success (see Supplementary Information). Countries with high and very high HDIs were more successful than low and medium HDI countries, owing to higher redundancy in management tactics and stronger central governance structures. Industrial fisheries scored higher than artisanal fisheries mainly because of stronger enforcement mechanisms, whereas inland fisheries were less successful than coastal and offshore fisheries owing mostly to weaker social capital and short-term co-management arrangements. Co-management systems thrived in benthic and demersal fisheries, especially when accompanied by protected areas, territorial user rights for fishing (TURFs) and community or individual quotas allocated to well-defined groups of fishers. In contrast, less successful co-management observed in multi-species fisheries could be related to a mismatch between scales of distribution and mobility of stocks and the area of influence of the
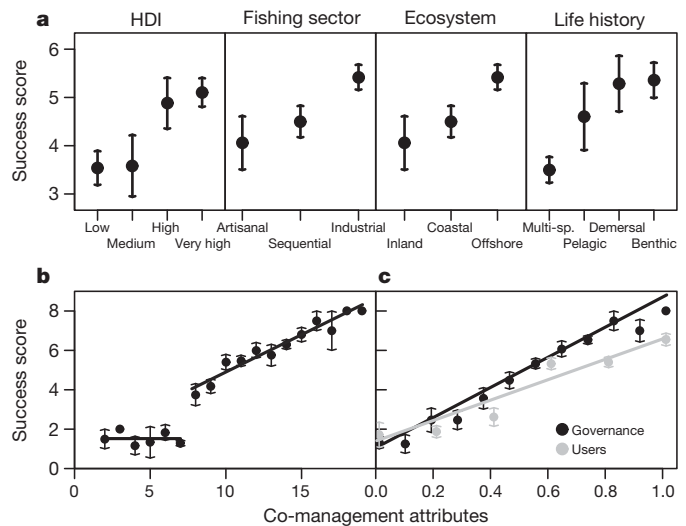

Figure 2 | Fisheries co-management performance. a, Success score discriminated by the HDI, fishing sector, ecosystem and life history. Multi-sp., multi-species. **b**, Success score correlated with the number of all co-management attributes present in the fishery. **c**, Success score correlated with proportion of governance and users' attributes separately (relative *x*-axis is shown for comparison purposes). Grouping variables are explained in Supplementary Table 2. All data are shown as mean ± s.e.m.

fishing process and the management system (Fig. 2a, Supplementary Fig. 3 and Supplementary Table 4).

There was a distinctive two-step pattern between success scores and the total number of attributes in each fishery. If fewer than eight attributes were present, the success score was close to zero, whereas above this threshold there was a strong positive relationship, with increasing attributes leading to higher success scores (Fig. 2b). Success scores were also more strongly correlated with the number of governance attributes present than with the number of users/community attributes (Fig. 2c and Supplementary Table 4). This indicates that even though co-management is enhanced by strong central governance systems, local community attributes were also necessary for success. These results demonstrate that the likelihood of co-management success increases when more management tools are added, providing redundancy in management regulations[19,20]. Further, no significant relationship ($P > 0.05$) was found between success and time frames of co-management regimes (omitting pre-implementation phase; mean ± standard deviation = 15.9 ± 9.8 years), indicating that failure or success is independent of the number of years the regime has been in place.

Using regression trees and random forests[21], we found that the most important co-management conditions necessary for successful management of fisheries are presence of community leaders, strong social cohesion, individual or community quotas, and community-based protected areas (Fig. 3a, b and Supplementary Table 2). Additional key attributes were enforcement mechanisms, long-term management policies and influence of fishers in local markets. Considering governance and users' attributes independently in the regression tree showed little differences in predictive accuracy compared to the joint tree (<4%) and between governance and users' trees (<5%). When analysed separately, community quotas were the most important management attribute followed by long-term management policies and protected areas, whereas leadership was by far the most significant users' attribute (Supplementary Fig. 4). These findings reinforce the notion that fisheries are complex social–ecological systems that need to be managed by addressing problems related not only to the resources themselves but to the people targeting them[22].

Leadership was critical for successful co-management of fisheries. Presence of at least one singular individual with entrepreneurial skills, highly motivated, respected as a local leader and making a personal
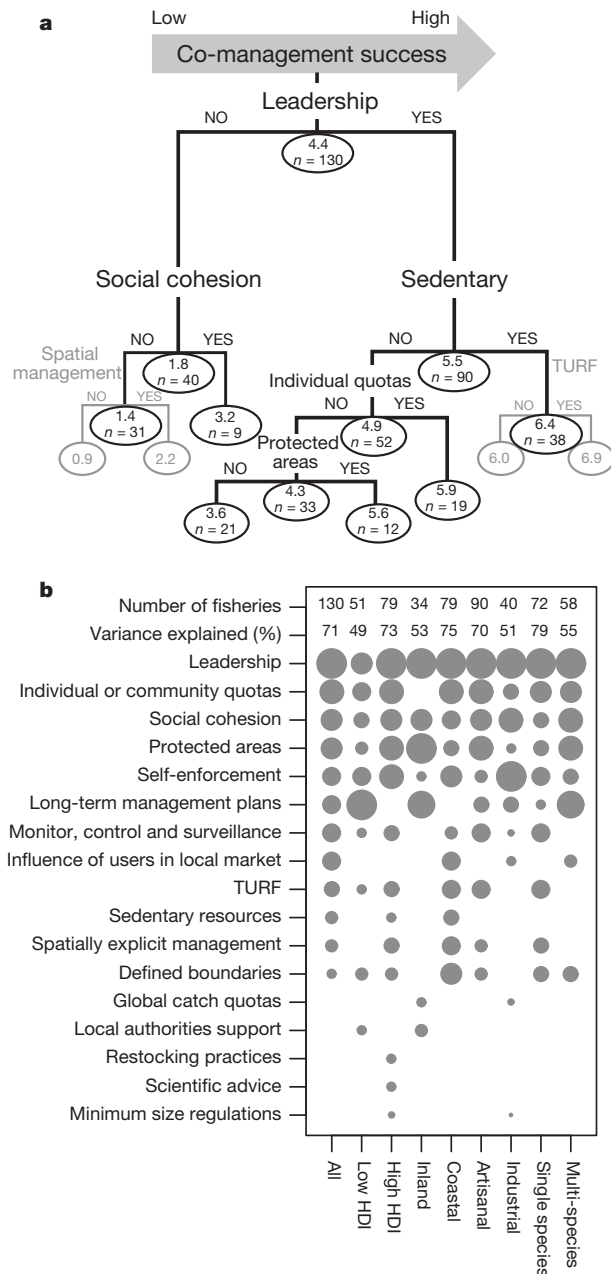
Figure 3 | Key co-management attributes for fisheries success. a, Regression tree showing the most important factors determining success. Higher branches offer greater explanatory power. Average success score and number of fisheries are listed at each node. The optimal tree explained 69% of the total deviance, and the vertical depth of each split is proportional to the variation explained by each attribute. **b**, Importance of individual attributes (rank proportional to circle size) for the full data set and for selected subsets of the data determined by random forests. The number of fisheries and variance explained are also indicated. Variables descriptions are given in Supplementary Table 2.

commitment to the co-management implementation process, was essential. Legitimate community leaders, when guided by collective interests and not self-benefits, give resilience to changes in governance, influence users' compliance to regulations and enhance conflict resolutions in quota allocations[23]. Community cohesion founded on norms, trust, communication, and connectedness in networks and groups was also an important global attribute leading to successful fisheries co-management. This robust social capital[7,24] serves as a buffer against changes in institutional arrangements, economic crises and resource overexploitation, and fosters sustainable co-management systems[3,25]. Our results show that additional resources should be spent on efforts to identify

community leaders and build social capital rather than only imposing management tactics without users' involvement.

Catch shares, both by individual or community quotas and by TURFs, were a key management condition towards co-management success. Well-designed and implemented catch shares have helped to prevent overfishing[26], promote stability[27] and ecological stewardship[28]. However, previous analyses of catch share programs have focused mainly on industrial fisheries in developed countries. We highlight the importance of users' security over catch or space in attaining social, economic and ecological success across all co-managed fisheries.

The effects of protected areas in achieving co-management success reaffirmed their strong link to social–ecological dynamics and the role of local communities in their successful implementation[29]. Their potential value for improving fisheries management depends on proper incentives, decentralized institutional arrangements and cohesive social organizations, all of which are more likely to happen under well-established co-management regimes. Spatial considerations, through clearly defined geographic boundaries (such as lake or enclosed bay) and sedentary life history of the resources contributed to co-management success by confining the number of users, lowering associated costs of information gathering, monitoring and enforcement, and restricting the spatial dynamics of fishing effort to well-defined areas.

Self-enforcing mechanisms contributed significantly to co-management success when guided by self-interests[24] (for example, through systems of penalties imposed by strong operational rules designed, enforced and controlled by local fishers). Influence of fishers in local markets characterized most accomplished co-management regimes, by allowing for specific marketing tactics, improved product quality, shorter intermediaries' chains, market timing coordination and eco-labelling strategies. This influence of users in local markets may result in multiple benefits to local communities, minimizing the probability of overexploitation and enhancing economic revenues by higher income per unit of effort[12].

Our study is, to our knowledge, the first comprehensive global assessment of social, economic and ecological attributes contributing to fisheries co-management success. Our synthesis shows that co-management holds great promise for successful and sustainable fisheries worldwide. However, there is an urgent need to gather long-term ecological, economic and social data from a variety of fisheries in a multidisciplinary context in order to compare empirically different degrees of users' involvement in management decisions and to better understand and improve fisheries co-management[30].

METHODS SUMMARY

We conducted a systematic search of the peer-reviewed and grey literature ($n = 1,168$ documents) to identify quantitative and qualitative evidences of the impacts of fisheries co-management practices around the world. We used the term community-based co-management to cover the whole spectrum of co-management arrangements (from formal consultation mechanisms between government and users to self-governance). The presence of well-established local co-management institutions with decision power in fisheries management was also used as compulsory criterion to classify a fishery as co-managed. Fisheries without sufficient or consistent information as well as co-management regimes in a pre-implementation phase were excluded from the analyses. For 130 fisheries (out of a total of 218 study cases; Supplementary Table 1) we compiled a database of 9 grouping or contextual variables including co-management type, co-management phase, duration of the management regime, HDI, Corruption Perception Index, Governance Index, ecosystem, fishing sector and resource type and 19 co-management attributes (Table 1 and Supplementary Table 2). We used aggregated social, economic and ecological binary outcomes to represent co-management success (success score; Supplementary Table 2). We built a regression tree model that graphically depicts quantitative relationships between predictor attributes and co-management success. Missing values were filled in using surrogate splits inside the regression model. A random forest model of 10,000 trees was used to estimate the relative importance of selected attributes in determining co-management success. The importance of contextual variables (for example, fishing sector) was also investigated by grouping them in the random forest models and by running independent models for each category (for example, artisanal, industrial). Model accuracy for trees and random forests were quantified using standard metrics,

and model selection was performed by backwards stepwise elimination of non-significant predictors (see Supplementary Information).

1. Food and Agriculture Organization of the United Nations. *FAO Yearbook: Fishery and Aquaculture Statistics* 2007 (FAO, 2009).
2. Worm, B. *et al.* Rebuilding global fisheries. *Science* **325,** 578–585 (2009).
3. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge Univ. Press, 1990).
4. Hardin, G. The tragedy of the commons. *Science* **162,** 1243–1248 (1968).
5. Costanza, R. *et al.* Principles for sustainable governance of the oceans. *Science* **281,** 198–199 (1998).
6. Dietz, T., Ostrom, E. & Stern, P. The struggle to govern the commons. *Science* **302,** 1907–1912 (2003).
7. Pretty, J. Social capital and the collective management of resources. *Science* **302,** 1912–1914 (2003).
8. Beddington, J. R., Agnew, D. J. & Clark, C. W. Current problems in the management of marine fisheries. *Science* **316,** 1713–1716 (2007).
9. Berkes, F. Community-based conservation in a globalized world. *Proc. Natl Acad. Sci. USA* **104,** 15188–15193 (2007).
10. Pomeroy, R. S. & Williams, M. J. *Fisheries Co-Management and Small-Scale Fisheries: A Policy Brief* (ICLARM, 1994).
11. Borrini-Feyerabend, G., Pimbert, M., Farvar, M. T., Kothari, A. & Renard, Y. *Sharing Power: Learning by Doing in Co-Management of Natural Resources Throughout the World* (IIED and IUCN/CEESP/CMWG, 2004).
12. Defeo, O. & Castilla, J. C. More than one bag for the world fishery crisis and keys for co-management successes in selected artisanal Latin American shellfisheries. *Rev. Fish Biol. Fish.* **15,** 265–283 (2005).
13. Wilson, J., Yan, L. & Wilson, C. The precursors of governance in the Maine lobster fishery. *Proc. Natl Acad. Sci. USA* **104,** 15212–15217 (2007).
14. Basurto, X. & Ostrom, E. The core challenges of moving beyond Garrett Hardin. *J. Nat. Resour. Pol. Res.* **1,** 255–259 (2009).
15. United Nations Development Programme. *The Human Development Index* (UNDP, 2009).
16. Ostrom, E. A general framework for analyzing sustainability of social-ecological systems. *Science* **325,** 419–422 (2009).
17. Hilborn, R. Defining success in fisheries and conflicts in objectives. *Mar. Policy* **31,** 153–158 (2007).
18. Chetty, R. K. *et al.* A systematic approach to preclinical and clinical safety biomarker qualification incorporating Bradford Hill's principles of causality association. *Clin. Pharmacol. Ther.* **88,** 260–262 (2010).
19. Castilla, J. C. & Defeo, O. Paradigm shifts needed for world fisheries. *Science* **309,** 1324–1325 (2005).
20. Stefansson, G. & Rosenberg, A. A. Combining control measures for more effective management of fisheries under uncertainty: quotas, effort limitation and protected areas. *Phil. Trans. R. Soc. B* **360,** 133–146 (2005).
21. Breiman, L. Random forests. *Mach. Learn.* **45,** 5–32 (2001).
22. Hilborn, R. Managing fisheries is managing people: what has been learned? *Fish Fish.* **8,** 285–296 (2007).
23. Olsson, P., Folke, C. & Hahn, T. Social-ecological transformation for ecosystem management: the development of adaptive co-management of a wetland landscape in southern Sweden. *Ecol. Soc.* **9,** 2 (2004).
24. Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **466,** 861–863 (2010).
25. Olsson, P., Folke, C. & Berkes, F. Adaptive co-management for building resilience in social–ecological systems. *Environ. Manage.* **34,** 75–90 (2004).
26. Costello, C., Gaines, S. D. & Lynham, J. Can catch shares prevent fisheries collapse? *Science* **321,** 1678–1681 (2008).
27. Essington, T. E. Ecological indicators display reduced variation in North American catch share fisheries. *Proc. Natl Acad. Sci. USA* **107,** 754–759 (2010).
28. Branch, T. A. How do individual transferable quotas affect marine ecosystems? *Fish Fish.* **10,** 39–57 (2009).
29. Pollnac, R. *et al.* Marine reserves as linked social–ecological systems. *Proc. Natl Acad. Sci. USA* (2010).
30. Levin, S. Crossing scales, crossing disciplines: collective motion and collective action in the Global Commons. *Phil. Trans. R. Soc. B* **365,** 13–18 (2010).

LETTER

Grains and grain boundaries in single-layer graphene atomic patchwork quilts

Pinshane Y. Huang[1]*, Carlos S. Ruiz-Vargas[1]*, Arend M. van der Zande[2]*, William S. Whitney[2], Mark P. Levendorf[3], Joshua W. Kevek[4], Shivank Garg[3], Jonathan S. Alden[1], Caleb J. Hustedt[5], Ye Zhu[1], Jiwoong Park[3,6], Paul L. McEuen[2,6] & David A. Muller[1,6]

The properties of polycrystalline materials are often dominated by the size of their grains and by the atomic structure of their grain boundaries. These effects should be especially pronounced in two-dimensional materials, where even a line defect can divide and disrupt a crystal. These issues take on practical significance in graphene, which is a hexagonal, two-dimensional crystal of carbon atoms. Single-atom-thick graphene sheets can now be produced by chemical vapour deposition[1–3] on scales of up to metres[4], making their polycrystallinity almost unavoidable. Theoretically, graphene grain boundaries are predicted to have distinct electronic[5–8], magnetic[9], chemical[10] and mechanical[11–13] properties that strongly depend on their atomic arrangement. Yet because of the five-order-of-magnitude size difference between grains and the atoms at grain boundaries, few experiments have fully explored the graphene grain structure. Here we use a combination of old and new transmission electron microscopy techniques to bridge these length scales. Using atomic-resolution imaging, we determine the location and identity of every atom at a grain boundary and find that different grains stitch together predominantly through pentagon–heptagon pairs. Rather than individually imaging the several billion atoms in each grain, we use diffraction-filtered imaging[14] to rapidly map the location, orientation and shape of several hundred grains and boundaries, where only a handful have been previously reported[15–19]. The resulting images reveal an unexpectedly small and intricate patchwork of grains connected by tilt boundaries. By correlating grain imaging with scanning probe and transport measurements, we show that these grain boundaries severely weaken the mechanical strength of graphene membranes but do not as drastically alter their electrical properties. These techniques open a new window for studies on the structure, properties and control of grains and grain boundaries in graphene and other two-dimensional materials.

Figure 1a shows a large array of the suspended, single-layer graphene membranes used in this study. We grew predominately single-layer graphene films on copper foils by chemical vapour deposition[1] (CVD) using three different growth recipes, which we refer to as growth methods A, B and C. Unless otherwise stated, all data were taken on graphene grown with method A, which was similar to the recipe reported in ref. 1. Methods B and C are slight variations: method B uses ultrapure copper foils[18] (99.999% pure rather than 99.8%) and method C uses a rapid thermal processor furnace (Methods). These films were transferred onto holey silicon nitride or Quantifoil transmission electron microscopy (TEM) grids using two different techniques (Methods and Supplementary Information). One key innovation over previous graphene TEM sample fabrication[20] was the gentle transfer of the graphene onto a TEM grid using a minimum of polymer support and baking the samples in air to remove the polymer without liquid solvents.

This produces large arrays of free-standing graphene sheets covering up to 90% of the TEM grid holes.

To characterize these membranes at the atomic scale, we used aberration-corrected annular dark-field scanning transmission electron microscopy (ADF-STEM), where a 60-keV, ångström-scale electron beam is scanned over the sample while the medium- to high-angle scattered electrons are collected. Keeping the electron beam voltage below the ~100-keV graphene damage threshold was necessary to limit beam-induced damage. Properly calibrated, this technique images the location and atomic number[21] of each atom and, along with TEM, has been used to study the lattice and atomic defects of graphene and boron nitrene[19,21–23]. Figure 1b shows an ADF-STEM image of the crystal lattice within a single graphene grain. Away from the grain boundaries, such regions are defect free.

In Fig. 1c, two graphene grains meet with a relative misorientation of 27°, forming a tilt boundary. Additional images of grain boundaries are shown in Supplementary Figs 2c and 3. As highlighted in Fig. 1d, the two crystals are stitched together by a series of pentagons, heptagons and distorted hexagons. The grain boundary is not straight, and the defects along the boundary are not periodic. Although the boundary dislocation
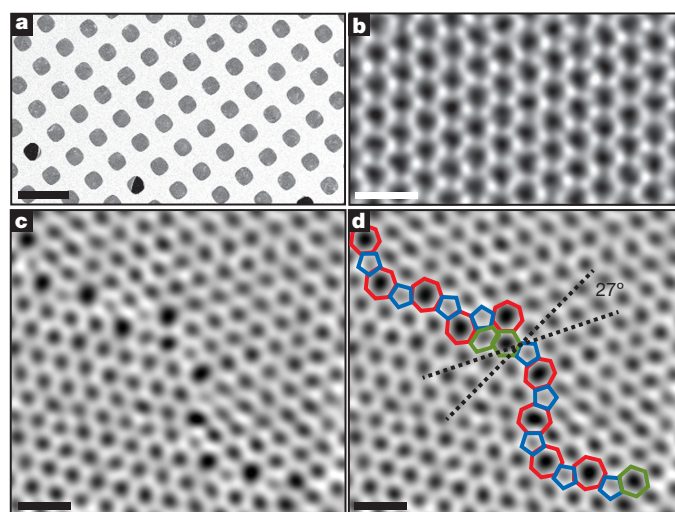

Figure 1 | Atomic-resolution ADF-STEM images of graphene crystals. a, Scanning electron microscope image of graphene transferred onto a TEM grid with over 90% coverage using novel, high-yield methods. Scale bar, 5 µm. **b,** ADF-STEM image showing the defect-free hexagonal lattice inside a graphene grain. **c,** Two grains (bottom left, top right) intersect with a 27° relative rotation. An aperiodic line of defects stitches the two grains together. **d,** The image from **c** with the pentagons (blue), heptagons (red) and distorted hexagons (green) of the grain boundary outlined. **b–d** were low-pass-filtered to remove noise; scale bars, 5 Å.

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA. [2]Department of Physics, Cornell University, Ithaca, New York 14853, USA. [3]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. [4]Department of Physics, Oregon State University, Corvallis, Oregon 97331, USA. [5]Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA. [6]Kavli Institute at Cornell for Nanoscale Science, Ithaca, New York 14853, USA.
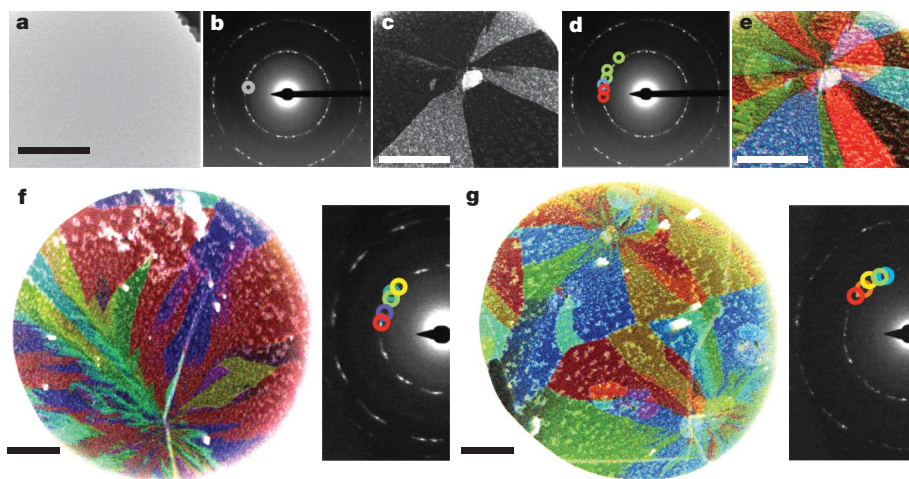*These authors contributed equally to this work.

Figure 2 | Large-scale grain imaging using DF-TEM. a–e, Grain imaging process. **a**, Samples appear uniform in bright-field TEM images. **b**, Diffraction pattern taken from a region in **a** reveals that this area is polycrystalline. Placing an aperture in the diffraction plane filters the scattered electrons forming **c**, a corresponding dark-field image showing the real-space shape of these grains. **d**, Using several different aperture locations and colour-coding them produces **e**, a false-colour, dark-field image overlay depicting the shapes and lattice orientations of several grains. **f, g**, Images of regions where many grains emanate from a few points. Scale bars, 500 nm.

resembles structures proposed theoretically[11,13], its aperiodicity contrasts with many of these models and will strongly affect the predicted properties of grain boundaries. By analysing atomic scattering intensities[21], we confirm that the boundary is composed entirely of carbon. In addition, although high electron beam doses could induce isolated bond rotations (Supplementary Fig. 3), the boundary was largely stable under the 60-keV electron beam. Thus, the polycrystalline graphene is a strongly bonded, continuous carbon membrane. We also note that many grain boundaries are decorated by lines of surface particles and adsorbates (Supplementary Fig. 4), suggesting that, as predicted[10], they may be more chemically reactive than the pristine graphene lattice.

Both STEM and TEM, which determine the positions and identities of atomic nuclei, and complementary scanning tunnelling microscopy, used to probe valence wavefunctions[15–17], are invaluable for understanding the local properties of grain boundaries. Using these atomic-resolution approaches, however, tens of billions to hundreds of billions of pixels would be needed to image even a single micrometre-scale grain fully, with estimated acquisition times of a day or more. Other candidates for characterizing grains on larger scales, such as low-energy electron microscopy[18] and Raman microscopy[3], typically cannot resolve small grains and may be difficult to interpret. Fortunately, electron microscopy offers an ideal technique for imaging grains on the necessary length scales: dark-field TEM (DF-TEM), which is a high-throughput, diffraction-sensitive imaging technique[14] that can be implemented on most TEMs built in the past sixty years. This method is usually applied to foils about 100–300-nm thick[14], but we demonstrate below that, remarkably, it also works on single-atom-thick sheets—even on samples too dirty for atomic-resolution imaging. In this manner, DF-TEM provides a nanometre- to micrometre-scale grain analysis that complements ADF-STEM to give a complete understanding of graphene grains on every relevant length scale.

Figure 2a, b shows a bright-field TEM image of a graphene sheet along with the selected-area electron diffraction pattern created from this region of the membrane. Owing to graphene's six-fold symmetry, electron diffraction from a single graphene crystal results in one set of six-fold-symmetric spots. Figure 2b contains many such families of spots, indicating that the field of view contains several grains of different orientations. DF-TEM images these grains one by one with few-nanometre resolution using an objective aperture filter in the back focal plane to collect electrons diffracted through a small range of angles, as shown by the circle in Fig. 2b. The resulting real-space image (Fig. 2c) shows only the grains corresponding to these selected in-plane lattice orientations and requires only a few seconds to acquire. By repeating this process using several different aperture filters, then colouring and overlaying these dark-field images (Fig. 2d, e), we create complete maps of the graphene grain structure, colour-coded by lattice orientation, as shown in Fig. 2e–g.

The images obtained are striking. The grains have complex shapes and many different crystal orientations. In Fig. 2e–g, we observe special locations from which many grains emanate. Small particles and multilayer graphene also are often found near these sites; see, for example, Fig. 2e, top right. Both the average spacing (2–4 μm) and shapes of these radiant sites when we use growth method A are comparable with Raman and scanning electron microscope observations of graphene nucleation[1,3], suggesting that these locations are probably nucleation sites. Similar structures have been observed in studies of crystallization in colloids and are consistent with crystallization around impurities[24]. Similar multigrain nucleation on copper has recently been observed using low-energy electron microscopy[18]. Significantly, each apparent nucleation site gives rise to many grains of different orientations, resulting in a mean grain size much smaller than the nucleation density.

The distributions of grain size and relative angular orientation are readily determined from DF-TEM images. As discussed below, grain sizes are dependent on growth conditions, here ranging from hundreds of nanometres to tens of micrometres for slight changes in growth conditions. In Fig. 3a, we plot a histogram of grain sizes across several samples grown using method A. The mean grain size, defined as the square root of the grain area, is 250 ± 11 nm (s.e.m.). This size is much smaller than the grain size of the copper substrate[1,4] (100 μm–1 mm) and typical lateral grains measured in bulk, highly ordered pyrolytic graphite[25] (6–30 μm). The inset in Fig. 3a shows the cumulative probability of finding multiple grains in a given area. This plot demonstrates that micrometre-scale CVD graphene devices produced from this set of films will nearly always contain multiple grains. Figure 3b shows a histogram of the relative crystallographic angles between adjacent grains. Because of graphene's six-fold crystal symmetry, the diffractive imaging technique only determines grain rotations modulo 60°. Consequently, the measurable difference between grain orientations is from 0 to 30° (with, for example, 31° measured as 29°). We observe a surprising and robust preference for low-angle (~7°) grain boundaries and high-angle (~30°) boundaries similar to that seen in Fig. 1.

Additional information about these orientations comes from the larger-area diffraction patterns in Fig. 3c, which we created by averaging diffraction data sampled across 1,200-μm² regions of graphene. The broadened diffraction peaks in Fig. 3c (left) show a distinct six-fold pattern, indicating that a significant fraction of the grains are approximately aligned across large areas This alignment can also be seen in Fig. 3d, which is a low-magnification DF-TEM image showing grains with a small (~10°) range of in-plane lattice orientations. Almost half of the membrane appears bright, indicating that these grains are all approximately aligned. In contrast, a dark-field image of randomly oriented grains would only show roughly one-sixth (10°/60°) of the graphene membrane. In the diffraction pattern of a separately grown sample (Fig. 3c, right), we instead find a clear 12-fold periodicity,
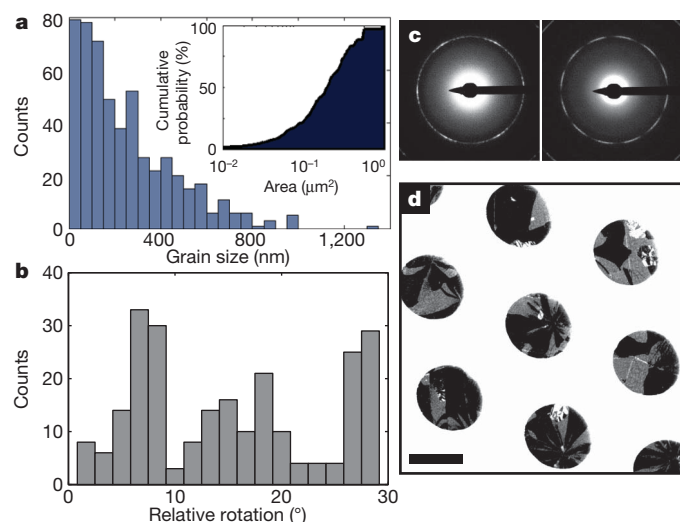
Figure 3 | Statistical analysis of grain size and orientation. a, Histogram of grain sizes, taken from three representative samples using DF-TEM. The mean grain size is 250 ± 11 nm (s.e.m., $n = 535$). Inset, plot of the cumulative probability of having more than one grain given the area of a device. **b,** Histogram of relative grain rotation angles measured from 238 grain boundaries. **c, d,** Large-area diffraction patterns (**c**) and a low-magnification DF-TEM image (**d**) show that grains are globally aligned near particular directions. Scale bar, 2 μm.

indicating that there are two main families of grains rotated from one another by 30°. These distributions, which often contain smaller sub-peaks (Supplementary Fig. 6), are consistent with the frequent observation of low-angle and high-angle (~30°) grain boundaries. We attribute these alignments to registry to the copper substrate used for graphene growth. Such registry has recently been observed in low-energy electron microscopy and scanning tunnelling microscopy studies of graphene growth on copper (100) and (111) surfaces[15,16,18].

By directly correlating grain structure with growth methods, these DF-TEM methods can be used to build on recent studies[3] that have demonstrated links between island nucleation density and growth conditions. Fig. 4a–c shows three composite DF-TEM images of graphene grown using methods A, B and C. The slight differences between growth methods effected significant changes in the grain size, shape and crystallographic orientation of the CVD graphene. For example, with growth method C we observed grains averaging 1–4 μm (Fig. 4c), which is an order of magnitude larger than the grains grown using method A. Our DF-TEM methods provide a powerful characterization tool for understanding and controlling grain growth, which will be a rich field of study important for graphene applications.

The ability to image the grain structure in graphene monolayers easily opens the door to the systematic exploration of the effects of grain structure on the physical, chemical, optical and electronic properties of

graphene membranes. We find that such studies are further facilitated because grain boundaries are visible in scanning electron microscopy and atomic force microscopy (AFM) phase imaging owing to preferential decoration of the grain boundaries with surface contamination (Fig. 5a and Supplementary Figs 9 and 10). Below, we show two examples probing the electrical and mechanical properties of grain boundaries.

We first examine the failure strength of the polycrystalline CVD graphene membranes (growth method A) using AFM. We used AFM phase imaging to image grains (Fig. 5a) and then pressed downwards with the AFM tip to test the mechanical strength of the membranes. As seen in Fig. 5b, the graphene tears along the grain boundaries. From repeated measurements, we find that failure occurs at loads of ~100 nN, which is an order of magnitude lower than typical fracture loads of 1.7 μN reported for single-crystal exfoliated graphene[26]. Thus, the strength of polycrystalline graphene is dominated by its grain boundaries.

We probed the electrical properties of polycrystalline graphene by fabricating electrically contacted devices using graphene from all three growth methods. Figure 4d shows a histogram of mobilities extracted from four-point transport measurements. Devices grown using methods A, B and C have room-temperature mobilities of $1,000 \pm 750$, $7,300 \pm 1,100$ and $5,300 \pm 2,300$ cm^2 V^{-1} s^{-1} (s.d.), respectively. The mobilities for growth method A are comparable to previous results on CVD graphene[1], whereas the mobilities of growth methods B and C are closer to those reported for exfoliated graphene[27] (1,000–20,000 cm^2 V^{-1} s^{-1}). By comparing these measurements with the corresponding DF-TEM images in Fig. 4a–c, we are surprised to find that, although mobility is clearly affected by growth conditions, high mobility does not directly correlate with large grain size.

To complement these bulk electrical measurements, we used scanning probe a.c. electrostatic force microscopy[28] (AC-EFM) to test the resistivity of individual grain boundaries. We fabricated suspended graphene membrane devices[29]. One of these is shown schematically in Fig. 5c, where we also plot the relative potential along a graphene membrane between two biased electrodes, measured using AC-EFM. In this plot, high-resistance grain boundaries would manifest as sharp drops in potential. The graphene in these devices (growth method A) had a mean grain size of 250 nm, so a line scan across these 3-μm-long membranes should cross an average of 12 grains. However, no noticeable potential drops were detected, indicating that most grain boundaries in these devices are not strongly resistive interfaces. By assuming that the grain boundary runs perpendicular to the line scan, we estimate an upper bound on the grain boundary resistance of $R_{GB} < 60\ \Omega\ \mu m/L$, where L is the length of the grain boundary, to be compared with the sheet resistance of $R_{graphene} = 700\ \Omega/\square$ for the entire device. In other words, the resistance of the grain boundaries is less than one-third the resistance of a 250-nm grain. Further measurements on six additional graphene membranes, both suspended and unsuspended, and from different growth methods, produced similar results. This small impact of grain boundaries stands in stark contrast to other materials, such as complex oxides, where a single grain boundary can lead to a million-fold increase in resistance over single crystals[30].
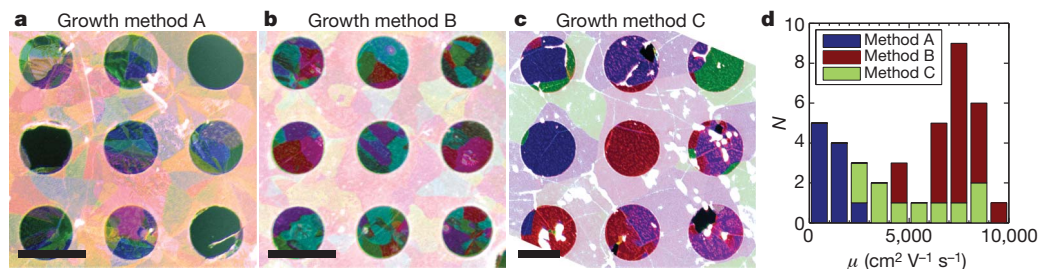

Figure 4 | Grain structure and mobilities for three growth conditions. a–c, Composite DF-TEM images of grain structure show variations with growth condition. The mean grain sizes are 250 ± 11 nm (s.e.m.; growth method A, 99.8% pure copper), 470 ± 36 nm (s.e.m.; growth method B, 99.999% pure (ultrapure) copper) and 1.7 ± 0.15 μm (s.e.m.; growth method C (rapid thermal

anneal)). The graphene is visible through the 20-nm, perforated amorphous-carbon Quantifoil support film. The graphene is broken over three of the perforations in **a**. Scale bars, 2 μm. **d,** Vertically stacked histogram of room-temperature mobilities, μ, measured from 39 devices using graphene growth methods A, B, and C. N, number of devices. See Methods for further details.
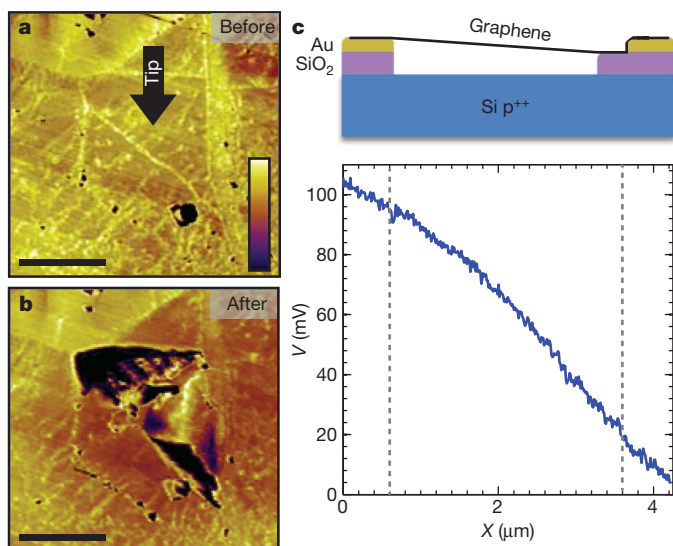
Figure 5 | AFM indentation and AC-EFM studies of graphene grain boundaries. a, b, AFM phase images of a graphene grain before and after an indentation measurement. **a**, Indentation takes place at the centre of this grain as shown by the arrow. **b**, The region is torn along grain boundaries after indentation. Scale bars, 200 nm. **c**, Electrostatic potential, averaged over three adjacent line scans along a suspended graphene sheet between two electrodes (schematic at top) and measured using AC-EFM. Although on average each line scan should cross 12 grains, no measureable features are present. Dashed lines indicate the locations of the electrodes.

The imaging techniques reported here provide the tools to characterize graphene grains and grain boundaries on all relevant length scales. These methods will be crucial both for exploring synthesis strategies to optimize grain properties and for studies, such as those described above, on the microscopic and macroscopic impact of grain structure on graphene membranes. Thus, these results represent a significant step forward in realizing the ultimate promise of atomic membranes in electronic, mechanical and energy-harvesting devices.

METHODS SUMMARY

TEM/STEM. We did ADF-STEM imaging using a NION UltraSTEM100 with imaging conditions similar to those used in ref. 21. At 60 kV, using a 33–35-mrad convergence angle, the electron probe was close to 1.3 Å in size and did not damage the graphene. Images presented in Figs 1–4 were acquired with the medium-angle annular dark-field detector with acquisition times of between 16 and 32 μs per pixel. For TEM imaging, we used a FEI Technai T12 operated at 80 kV. Acquisition times for dark-field images were 5–10 s per frame. The spatial resolution in dark-field images ranges from 1 to 10 nm and is set by the size of the objective filtering aperture, in a trade-off between real-space resolution and angular resolution in reciprocal space.

Scanning probe measurements. For AFM deflection measurements, we used a MFP3D scope from Asylum Research. We used silicon AFM probes (Multi75Al, Budget Sensors) with a resonant frequency of ~75 kHz, a force constant of ~3 N m^{-1} and a tip radius of <10 nm. All imaging was done in tapping mode. For AC-EFM measurements, a DI 4100 AFM with a signal access module was operated in lift mode with a constant probe tip voltage, $V_{tip} = 2$ V, a lift height of 10 nm and no piezo drive on the tip. An a.c. voltage of $V_0 = 1$ V was applied through the electrodes at the resonance frequency of the EFM cantilever, $f_{cant} \approx 77$ kHz. An electrostatic force drives the EFM cantilever to resonate, and the amplitude of motion is measured.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Li, X. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324,** 1312–1314 (2009).
2. Reina, A. *et al.* Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition. *Nano Lett.* **9,** 30–35 (2009).
3. Li, X. *et al.* Graphene films with large domain size by a two-step chemical vapor deposition process. *Nano Lett.* **10,** 4328–4334 (2010).
4. Bae, S. *et al.* Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnol.* **5,** 574–578 (2010).
5. Cervenka, J. & Flipse, C. F. J. Structural and electronic properties of grain boundaries in graphite: planes of periodically distributed point defects. *Phys. Rev. B* **79,** 195429 (2009).
6. Peres, N. M. R., Guinea, F. & Castro-Neto, A. H. Electronic properties of disordered two-dimensional carbon. *Phys. Rev. B* **73,** 125411 (2006).
7. Yazyev, O. V. & Louie, S. G. Electronic transport in polycrystalline graphene. *Nature Mater.* **6,** 806–809 (2010).
8. Mesaros, A., Papanikolaou, S., Flipse, C. F. J., Sadri, D. & Zaanen, J. Electronic states of graphene grain boundaries. *Phys. Rev. B* **82,** 205119 (2010).
9. Cervenka, J., Katsnelson, M. I. & Flipse, C. F. J. Room-temperature ferromagnetism in graphite driven by two-dimensional networks of point defects. *Nature Phys.* **5,** 840–844 (2009).
10. Malola, S., Hakkinen, H. & Koskinen, P. Structural, chemical, and dynamical trends in graphene grain boundaries. *Phys. Rev. B* **81,** 165447 (2010).
11. Liu, Y. & Yakobson, B. I. Cones, pringles, and grain boundary landscapes in graphene topology. *Nano Lett.* **10,** 2178–2183 (2010).
12. Grantab, R., Shenoy, V. B. & Ruoff, R. S. Anomalous strength characteristics of tilt grain boundaries in graphene. *Science* **330,** 946–948 (2010).
13. Yazyev, O. V. & Louie, S. G. Topological defects in graphene: dislocations and grain boundaries. *Phys. Rev. B* **81,** 195420 (2010).
14. Hirsch, P., Howie, A., Nicholson, R., Pashley, D. W. & Whelan, M. J. *Electron Microscopy of Thin Crystals* (Krieger, 1965).
15. Zhao, L. *et al.* The atomic-scale growth of large-area monolayer graphene on single-crystal copper substrates. Preprint at ⟨http://arxiv.org/abs/1008.3542⟩ (2010).
16. Gao, L., Guest, J. R. & Guisinger, N. P. Epitaxial graphene on Cu(111). *Nano Lett.* **10,** 3512–3516 (2010).
17. Cockayne, E. *et al.* Rotational grain boundaries in graphene. Preprint at ⟨http://arxiv.org/abs/1008.3574⟩ (2010).
18. Wofford, J. M., Nie, S., McCarty, K. F., Bartlet, N. C. & Dubon, O. D. Graphene islands on Cu foils: the interplay between shape, orientation, and defects. *Nano Lett.* **10,** 4890–4896 (2010).
19. Park, H. J., Meyer, J., Roth, S. & Skakalova, V. Growth and properties of few-layer graphene prepared by chemical vapor deposition. *Carbon* **48,** 1088–1094 (2010).
20. Regan, W. *et al.* A direct transfer of layer-area graphene. *Appl. Phys. Lett.* **96,** 113102 (2010).
21. Krivanek, O. L. *et al.* Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* **464,** 571–574 (2010).
22. Hashimoto, A., Suenaga, K., Gloter, A., Urita, K. & Iijima, S. Direct evidence for atomic defects in graphene layers. *Nature* **430,** 870–873 (2004).
23. Meyer, J. C. *et al.* Direct imaging of lattice atoms and topological defects in graphene membranes. *Nano Lett.* **8,** 3582–3586 (2008).
24. de Villeneuve, V. W. A. *et al.* Hard sphere crystal nucleation and growth near large spherical impurities. *J. Phys. Condens. Matter* **17,** S3371–S3378 (2005).
25. Park, S., Floresca, H. C., Suh, Y. & Kim, M. J. Electron microscopy analyses of natural and highly oriented pyrolytic graphites and the mechanically exfoliated graphenes produced from them. *Carbon* **48,** 797–804 (2010).
26. Lee, C., Wei, X., Kysar, J. W. & Hone, J. Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science* **321,** 385–388 (2008).
27. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6,** 183–191 (2007).
28. Bachtold, A. *et al.* Scanned probe microscopy of electronic transport in carbon nanotubes. *Phys. Rev. Lett.* **84,** 6082–6085 (2000).
29. Van Der Zande, A. M. *et al.* Large-scale arrays of single-layer graphene resonators. *Nano Lett.* **10,** 4869–4873 (2010).
30. Thiel, S. *et al.* Electron scattering at dislocations in LaAlO3/SrTiO3 interfaces. *Phys. Rev. Lett.* **102,** 046809 (2009).

Author Contributions P.Y.H., C.S.R.-V. and A.M.v.d.Z. contributed equally to this work. Electron microscopy and data analysis were carried out by P.Y.H. and D.A.M., with Y.Z. contributing to initial DF-TEM. Graphene growth and sample fabrication were done by A.M.v.d.Z. and C.S.R.-V. under the supervision of P.L.M. and J.P., aided by M.P.L., S.G., W.S.W., J.W.K., J.S.A. and C.J.H. AC-EFM, mobility measurements and analysis were done by A.M.v.d.Z. and P.L.M., aided by C.S.R.-V. and J.W.K. AFM mechanical testing and analysis were done by C.S.R.-V. and J.P., aided by S.G. All authors discussed the results and implications at all stages. P.Y.H., A.M.v.d.Z., C.S.R.-V., P.L.M., J.P. and D.A.M. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.A.M. (david.a.muller@cornell.edu).

METHODS

ADF-STEM. ADF-STEM imaging was conducted using a NION UltraSTEM100 operated at 60 kV. Imaging conditions were similar to those used in ref. 21. Using a 33–35-mrad convergence angle, our probe size was close to 1.3 Å. Because the low-voltage electron beam was below the damage threshold energy[31], the pristine graphene lattice remains stable and defect free. High electron beam doses could induce isolated bond rotations at grain boundaries (Supplementary Fig. 4) similar to those seen in ref. 32. Images presented in Figs 1–4 were acquired with the medium-angle annular dark-field detector with acquisition times of between 16 and 32 μs per pixel.

DF-TEM. TEM imaging was conducted using a FEI Technai T12 operated at 80 kV, which did not cause any apparent damage to the graphene membranes. Acquisition time for dark-field images were 5–10 s per frame. The spatial resolution in dark-field images ranges from 1 to 10 nm and is set by the size of the objective filtering aperture in a trade-off between real-space resolution and angular resolution in reciprocal space.

AC-EFM. A DI 4100 AFM with a signal access module was operated in lift mode with tip voltage $V_{tip} = 2$ V, a lift height of 10 nm and no piezo drive on the tip. An a.c. voltage $V_0 = 1$ V was applied through the electrodes at the resonance frequency of the EFM cantilever, $f_{cant} \approx 77$ kHz. An electrostatic force drives the EFM cantilever to resonate, and the amplitude of motion is measured.

AFM imaging and deflection measurements. For AFM deflection measurements, we used a MFP3D scope from Asylum Research. We used silicon AFM probes (Multi75Al, Budget Sensors) with a resonant frequency of ~75 kHz, a force constant of ~3 N m^{-1} and a tip radius of <10 nm. All imaging was done in tapping mode. Images were taken with resolutions of 512×512 or $1,024 \times 1,024$, with acquisition times of at most 10 min.

Graphene growth. We grew single-layer graphene using CVD on copper foils in three ways. Growth method A: similar to methods described in ref. 1, we annealed a 99.8% pure copper foil (Alfa Aesar #13382) at 1,000 °C at low pressure with an H$_2$ flow of 7 standard cubic centimetres per minute (s.c.c.m.) for 10 min. We then grew the graphene at 1,000 °C by flowing CH$_4$:H$_2$ at 150:7 s.c.c.m. for 10–15 min (varying growth time within this range did not yield noticeably different results). Samples are cooled for ~50 min while the CH$_4$:H$_2$ flow is maintained. Growth method B: this is identical to method A, except we used higher purity (99.999%) copper foil (Alfa Aesar #10950). Growth method C: we used a rapid thermal processor tube furnace with a ~4″ inner diameter (MTI Corporation). We annealed copper foil (99.8% purity) at 1,000 °C (H$_2$, 300 s.c.c.m.) for 30 min, and then grew the graphene at 1,000 °C (CH$_4$:H$_2$, 875:300 s.c.c.m.) for 60 min.

Samples for DF-TEM. We transferred the graphene either to commercial holey SiN TEM grids (such as PELCO Holey Silicon Nitride Support Films) with 2.5-μm-diameter holes or to Quantifoil holey carbon TEM grids to allow imaging of larger grains. Quantifoil grids are typically 10–20 nm thick, which is thin enough to allow DF-TEM imaging through the carbon support.

The fabrication for DF-TEM samples is a gentle graphene transfer method using a thin PMMA support, which produced roughly 90% coverage of TEM grid holes (that is, 90% of grid holes were uniformly covered with suspended graphene). After graphene growth on a copper foil, a thin layer of PMMA was spun onto the graphene (2% in anisole, 4,000 r.p.m. for 30 s), without a post-baking step. Copper was then etched away by floating the foil, PMMA side up, in a HCl/FeCl$_3$ copper etchant (Transene, Type 100/200). Next, the graphene and polymer support were washed by transferring them to deionized-water baths, taking care to not bring the PMMA into contact with liquids, to avoid depositing unwanted residues on the PMMA side of this layer. Finally, the PMMA–graphene layer is scooped out in pieces onto TEM grids. PMMA can be thermally decomposed[33], which is a gentler process than using liquid solvent rinses. We baked our samples in air (350 °C for 3–4 h), without the use of an argon flow, which can slow the cleaning effect substantially. This step removes the PMMA layer, leaving the graphene freely suspended in a liquid-free release process. These high-yield samples were used in DF-TEM because they provided enough clean graphene to image large numbers of grains.

Samples for ADF-STEM. Our secondary technique produced cleaner, but lower-yield, graphene using a polymer-free transfer method. This technique is similar to the methods of ref. 20, in which TEM grids are placed on top of the foil before etching and attached by dropping methanol on the grids. Our main addition to this technique was to bake the final samples in a series of annealing processes increasing in temperature. The grids were then baked in air at 350 °C for 2 h. In this method, the samples are annealed in ultrahigh vacuum by ramping the temperature to 950 °C, holding this temperature steady for 15 min and then cooling to room temperature without active cooling. This annealing is done below the graphene growth temperature, and the micrometre-scale grain structure did not change afterwards. Thus, any change that may result from annealing should be small in comparison with changes occurring during the formation of the grain boundaries. A final step was to anneal the grids at 130 °C for >8 h before transferring them in air to the TEM. Because this transfer method uses no support film for the graphene as it is transferred, this method was a comparatively low-yield transfer process with coverage of just a few per cent over the holes. The advantage to this technique over the polymer-based transfer is that it produced graphene with less surface carbon contamination—regions hundreds of nanometres wide appeared atomically clean in ADF-STEM images.

Electrically contacted samples. We fabricated top-gated graphene devices in four-point probe geometry (shown in Supplementary Fig. 11a, b, with electrodes labelled). A transferred graphene film was patterned by photolithography and a 10-s exposure to an oxygen plasma to define the graphene strips. This was followed by fabricating 1.5-nm Ti/4.5-nm Au electrodes. We patterned a top gate, to measure the charge mobility in graphene, by electron beam evaporation first of 90 nm of silicon oxide as a dielectric layer and then of a Cr/Au layer (1.5 nm/50 nm), without breaking vacuum between each evaporation.

For the EFM measurements, we fabricated electrically contacted, suspended graphene by growing single-layer graphene on copper using CVD; patterning the graphene into 3-μm-wide strips while still on the copper foil, using contact lithography; and transferring the patterned strips onto a substrate with pre-patterned gold electrodes and trenches.

31. Meyer, J. C., Chuvilin, A. & Kaiser, U. in *MC2009, Vol. 3: Materials Science* (eds Grogger, W., Hofer, F. & Polt, P.) 347–348 (Graz Univ. Technology, 2009).
32. Suenaga, K. *et al.* Imaging active topological defects in carbon nanotubes. *Nature Nanotechnol.* **2,** 358–360 (2007).
33. Jiao, L. *et al.* Creation of nanostructures with poly(methyl methacrylate)-mediated nanotransfer printing. *J. Am. Chem. Soc.* **130,** 12612–12613 (2008).